Université du Québec

INRS – Institut Armand-Frappier

FACTORS INVOLVED IN ALTERNATIVE POLYADENYLATION DURING HERPES SIMPLEX VIRUS 1 INFECTION.

by

Huda Hyjazie

Thesis presented

for the degree Masters in Science

in virology and immunology

Evaluation Jury

Jury president and internal examiner :	Patrick Labonté, PhD. INRS – Institut Armand-Frappier
External examiner :	François Dragon, PhD. Département des Sciences Biologiques UQAM
Research director :	Angela Pearson, PhD. INRS – Institut Armand-Frappier

Abstract

Herpes simplex virus 1 (HSV-1) is an important human pathogen that infects more than 80% of the population and causes serious disease in immunocompromised individuals and newborns. This virus has a dense DNA genome that codes for more than 80 genes, thus necessitating effective genetic regulation. One of the modes of regulation utilized by HSV-1 is alternative polyadenylation: a post-transcriptional event that allows one gene to use two different polyadenylation signals (PAS). This process generates transcripts that differ in their 3' untranslated region (3'UTR), which can affect translation efficiency among other things. Several HSV-1 genes including *UL24* and *UL38* are regulated by alternative polyadenylation. *UL24* short transcripts, which arise from the use of the *UL24* polyadenylation signal, are expressed with early kinetics, while *UL24* long transcripts, which arise from use of the *UL26* polyadenylation signal, are expressed with late kinetics.

We hypothesized that the switch in PAS usage is due to changes in the steady-levels of polyadenylation factors over the course of infection. We also hypothesized that the degree of association of the polyadenylation factors with the different polyadenylation sites changes over the course of the infection.

In order to test the first hypothesis, Vero and HeLa cells were either mock-infected or infected for 3, 6, 9, 12 and 16hrs, and cell lysates were prepared. We evaluated the steady-state levels of three subunits of polyadenylation factors, CPSF-100, CstF-64 and CFI_m-25, by Western blot. We did not observe any reproducible change in the levels of these factors for the first 12 hours of infection, and therefore, a change in their levels could not explain the shift in PAS usage of the alternatively polyadenylated *UL24* gene observed at around 10hpi. Several subunits could not be evaluated using this technique due to commercial antibody availability and specificity. Nevertheless, though we did not observe a reproducible change in the steady-state levels of the factors tested during the first 12 hours of infection, we cannot rule out the possibility that the localization and the RNA or protein interactions of the various factors or subunits may have been affected.

To test the second hypothesis, we had as an objective to identify proteins that interact with the 3'UTRs of alternatively polyadenylated transcripts during the infection. To identify these proteins, we developed an affinity purification system based on the S1 RNA aptamer. We evaluated lysates from mock-infected cells as well as cells infected for 5hrs and 10hrs. Bound proteins were analysed by liquid chromatography–mass spectrometry (LC-MS). Among other

proteins, we identified eEF1A and NF45, both nuleic acid binding proteins. eEF1A plays a role in translation fidelity and also assists in the nuclear export of mature tRNA, and NF45 is a transcriptional activator. These results serve as proof of principal for this strategy to identify proteins that interact with the 3'UTR of viral transcripts. Future work will be aimed at validating the proteins identified, and investigating other possible mechanisms involved in alternative polyadenylation such as post-translational modifications of polyadenylation factors. The results from this project will contribute to a better understanding of genetic regulation by alternative polyadenylation.

Huda Hyjazie

Angela Pearson

Acknowledgements

I would first like to thank my research director for overseeing this project. I want to thank her for her encouragement, enthusiasm and understanding, especially during the difficult times. It has been a pleasure being a member of her lab.

I would also like to thank my colleagues, present and past, which have always been pleasant and supportive: Dr. Nawel ben Abdeljelil, Luc Bertrand, Alain Boulende, Amélie Bourget, Dr. Maria Lymberopoulos, Dr. Carla Mota, Gabriel Ouellette, Dr. Annie Rochette, Pierre-Alexandre Rochette and Carolina Sanabria. I would like to give special thanks to Dr. Nawel ben Abdeljelil for her help in writing the French summary of this thesis and her availability to answer my frequent questions. I would also like to give special thanks to Dr. Maria Lymberopoulos, Pierre-Alexandre Rochette and Luc Bertrand, our "senior students" who were very present and helpful in the lab. Finally, I would like to express my gratitude to our dear and thoughtful technician Ginette Denis who gave our lab a special touch that we all appreciate so much. I would also like to thank Romain Grangeon from our neighboring lab for his entertaining sense of humour.

I would also like to thank Anastasia Nikolakakis for analysing our MS samples at the INRS-Institut-Armand-Frappier MS facilities.

Last but not least, I would like to thank my family for their support and encouragement throughout this Master's Degree and through all difficult times. I would especially like to thank my father Fayez Hyjazie, my brother Jihad Hyjazie, my sister Dua Hyjazie and my two aunts Fadia and Shadia Hijazie. Finally, I would not have been able to submit this work within the scheduled time of arrival of our baby, Bayaan Fatima Mohammad, if it were not for the help, support and encouragement of my dear husband Mubarak Sami Mohammad.

All in all, it was a great pleasure to work for this degree.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3'UTR	3' untranslated region
$CFII_m$	Mammalian cleavage factor II
CFI_m	Mammalian cleavage factor I
CPSF	Cleavage and polyadenylation specificity factor
CstF	Cleavage stimulation factor
CTD	C-terminal domain
DMEM	Dulbecco's Modified Eagle's Media
dsDNA	Double stranded DNA
DSE	Downstream element
E	Early
eEF1A	Translation elongation factor 1A
ELL	Elongation factor RNA polymerase
gB	Glycoprotein B
gC	Glycoprotein C
GST	Glutathione S-transferase
HCF	Host cell factor
HCMV	Human cytomegalovirus
HSV-1	Herpes simplex virus 1
IE	Immediate-early
L	Late
LC-MS	Liquid chromatography - mass spectrometry
LRRK2	Leucine-rich repeat serine/threonine-protein kinase 2
NF45	Nuclear factor 45
Oct-1	Octamer transcription factor 1
PABP	PolyA binding protein
PAP	PolyA polymerase
PAS	Polyadenylation signal
RNA pol II	RNA polymerase II
RRM	RNA recognition motif
U_L	Unique long
U_S	Unique short
vhs	Virion host shut-off
VZV	Varicella-Zoster virus

CHAPTER 1: INTRODUCTION

1.1 Herpes simplex virus I

1.1.1 Herpesviruses

The order of the *Herpesvirales* contains three families: *Herpesviridae*, *Alloherpesviridae* and *Malacoherpesviridae*. The *Herpesviridae* family is divided into three subfamilies: the *alphaherpesvirinae*, the *betaherpesvirinae* and the *gammaherpesvirinae* subfamilies (Davison et al., 2009). The *alphaherpesvirinae* are neurotropic viruses, with a short replication cycle, and infect primarily mucoepithelial cells. The *betaherpesvirinae* target mainly lymphocytes and have a longer replication cycle. The *gammaherpesvirinae* replicate in epithelial cells and generally establish long-term latency in lymphocytes (Pellett and Roizman, 2007).

Herpes simplex virus I (HSV-1) is part of the *alphaherpesvirinae* subfamily, which also contains the human viruses HSV-2 and varicella-zoster virus (VZV), and is used as a prototype to study this family. The morphological appearance of HSV-1 and HSV-2 is virtually identical (Kieff et al., 1972). They are known to cause recurrent vesicular skin lesions. These lesions are recurrent since HSV-1 and HSV-2 are neurotropic viruses that establish latency in the sensory ganglia and are capable of reactivation (reviewed in Schleiss, 2009).

1.1.2 Pathogenesis

In the United States, 18-35% of children and approximately 90% of adults are infected with HSV-1 (reviewed in Desselberger, 1998). Also, approximately 1,500 newborns are infected with HSV each year (reviewed in Kimberlin, 2004). Close personal contact with an infected individual leads to a primary infection, which begins with an incubation period of about four days. Virus shedding in mouth and stool occurs for seven to ten days, and neutralizing antibodies appear four to seven days after clinical onset, and peak about two weeks later (reviewed in Jacobs, Breakefield, and Fraefel, 1999). Because of its neurotropism, HSV-1 is also being exploited for the development of gene therapy vectors (Epstein, 2009).

The life cycle of HSV involves both lytic (productive) and latent (non-productive) infection. Upon entry at a mucosal surface or at a break in the skin, HSV infects epithelial cells and undergoes productive infection and production of progeny virions that spread to infect neighbouring epithelial cells. The virus produced also infects innervating sensory neurons by fusion with the neuronal membrane at the axonal termini. Through retrograde transport, the nucleocapsid is then transported to the nucleus in the cell body of the neuron in a ganglion, where the viral DNA is released and persists in the form of an episome, and where lytic expression is repressed (reviewed in Knipe and Cliffe, 2008). Latency by HSV-1 and HSV-2 is typically established in

neurons of trigeminal and dorsal root ganglia respectively. Reactivation of latent viral DNA caused by exposure to UV radiation, stress or trauma (Desselberger, 1998) leads to viral lytic gene expression; capsids generated are then transported to the axonal termini resulting in recurrent infections and virus shedding (reviewed in Knipe and Cliffe, 2008).

Virus replication at the site of infection causes a localized ulcerative lesion leading to edema, usually in the oropharyngeal mucosa. HSV-1 replication is not limited to one site; HSV-1 keratoconjunctivitis is the leading infectious cause of corneal blindness in developed countries, and in immunocompromised hosts, virus dissemination can lead to infection of the skin, the respiratory tract, the esophagus and the gastrointestinal tract. A more rare yet serious phenomenon, not limited to immunocompromised individuals and affecting 1,500 patients per year in the US (reviewed in Knipe and Cliffe, 2008), is the transport of virions to the central nervous system, where hemorrahagic encephalitis of the temporal lobe takes place and is life-threatening (reviewed in Jacobs, Breakefield, and Fraefel, 1999).

1.1.3 Genome

HSV-1 has a linear, double-stranded DNA (dsDNA) genome of 152kbp (Pellett and Roizman, 2007). It is composed of a unique long (U_L) region and a unique short (U_S) region (Wadsworth, Jacob, and Roizman, 1975), both of which are flanked by the inverted repeats. Present in equal quantities during infection are virions with one of four possible isoforms of the genome (Fig. 1); uninverted U_S and U_L (orientation arbitrarily chosen), inverted U_S and U_L, inverted U_S only and inverted U_L only (reviewed in Roizman, 1979; Sheldrick and Berthelot, 1975). The HSV-1 genome contains two origins of replication: ori_L, located in the U_L region (Weller et al., 1985), and ori_S located in the U_S region (Stow and McMonagle, 1983). The genome contains 84 single copy ORFs (Roizman, Knipe, and Whitley, 2007) which are oriented in downstream as well as in upstream direction, and there are overlapping reading frames, indicating a very high density of genetic information (reviewed in Rajcani, Andrea, and Ingeborg, 2004). The genes are separated into three categories based on their transcription kinetics, which are tightly regulated in a temporal cascade of immediate-early (IE), early (E) and late (L) genes (Honess and Roizman, 1974).

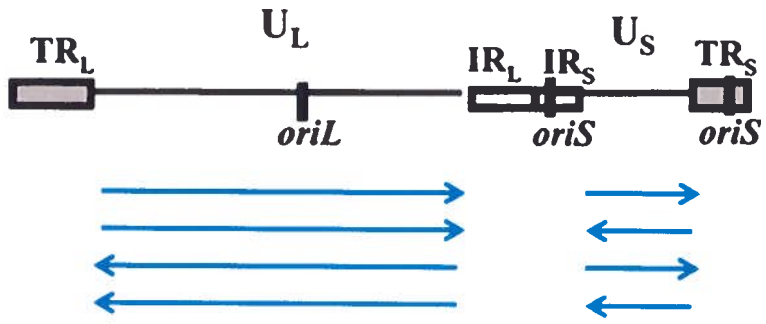

Fig. 1. The HSV-1 genome. Diagram showing the unique long U_L and unique short U_S segments of the HSV-1 genome, each flanked by their terminal repeat regions (TR) or their inverted internal repeat regions (IR). Also indicated are the origins of viral DNA replication (ori). The blue arrows indicate relative orientations of the U_L and U_S segments in the four different isoforms of the HSV-1 genome. (Modified from Sandri-Goldin, 2003)

1.1.4 Virion structure

The diameter of the virion is between 150 to 200 nm and consists of an envelope, the tegument, the capsid and the core which contains the virus genome (reviewed in Jacobs, Breakefield, and Fraefel, 1999) (Fig. 2). The envelope, which consists of a lipid membrane, is densely populated with glycoproteins (Wildy and Watson, 1962), and has a trilaminar appearance. The tegument is a partially ordered structure of proteins that surrounds the capsid tightly (reviewed in Mettenleiter, 2006). It includes the virus host shut-off protein (vhs) which degrades cellular mRNAs early during infection (reviewed in Jacobs, Breakefield, and Fraefel, 1999). It also includes VP16, which stimulates efficient transcription of the first set of viral genes expressed, the IE genes (Campbell, Palfreyman, and Preston, 1984; Post, Mackem, and Roizman, 1981). The capsid is made up of 162 capsomers, including 150 hexavalent and 12 pentavalent capsomers, and is characteristic of all herpesviruses (Booy et al., 1996). There are four essential capsid proteins: the major capsid protein UL19, the triplex monomer and dimer proteins UL38 and UL19 respectively, and the small capsomer-interacting protein UL35 (reviewed in Pellett and Roizman, 2007). The core of a mature virion contains the viral DNA, which is present in the shape of a toroid (Furlong, Swift, and Roizman, 1972).

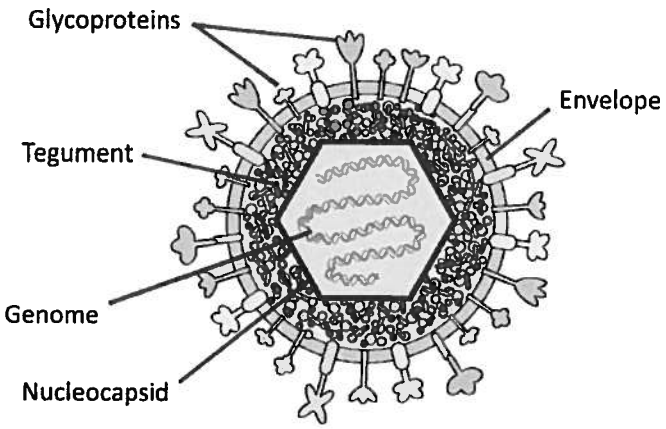

Fig. 2. Components of the HSV-1 virion. Representation of the virion components: the genome as dsDNA, the icosahedral nucleocapsid, the tegument proteins, the lipid envelope and the surface glycoproteins. (Modified from http://www.twiv.tv/virus-structure)

1.1.5 Virus lytic replication cycle

There are many steps involved in a productive HSV-1 infection (reviewed in Knipe and Cliffe, 2008) (Fig. 3). The virus first attaches to a cell through low affinity interaction between the glycoproteins C (gC) and B (gB) with cell-surface heparin sulphate proteoglycan (Herold et al., 1994; Herold et al., 1991). Recently, non-muscle myosin IIA was also identified as an entry receptor for gB (Arii et al., 2010). For virus entry, fusion with cellular membranes takes places through the concerted action of at least four glycoproteins, gD, gB, and the gH/gL complex (Turner et al., 1998). During virus entry, gD binds one of its receptors, either herpes virus entry mediator (HVEM) or nectin-1 (reviewed in Spear, Eisenberg, and Cohen, 2000). Receptor binding causes conformational changes to gD (Harrison, 2008) which enable it to activate gH/gL. Activation of the gH/gL heterodimer allows for the activation of gB's fusogenic activity (Atanasiu et al., 2010). The envelope then fuses with the plasma membrane. Still, in some cell types, pH-dependant fusion has also been observed (Heldwein and Krummenacher, 2008).

After fusion, the capsid is transported to the nuclear pores through active transport along the cellular microtubular cytoskeleton (Sodeik, Ebersold, and Helenius, 1997). The tegument protein UL14 enhances the nuclear localization of the viral regulatory protein VP16. It also plays a role in the efficient nuclear targeting of the capsid (Yamauchi et al., 2008). The capsid then binds the nuclear pores and the viral genome is released into the nucleus (McClain and Fuller, 1994). This release is thought to be facilitated by the tegument proteins (Newcomb and Brown,

1994). Once in the nucleus the genome is rapidly circularized (Garber, Beverley, and Coen, 1993). The genome is then transcribed by the cellular DNA-dependant RNA polymerase II (RNA pol II), beginning with the expression of the IE genes followed by the expression of the E genes. The nucleus undergoes reorganization to form replication compartments for viral DNA synthesis by a viral DNA-dependent DNA polymerase (Quinlan, Chen, and Knipe, 1984). The early protein ICP8, among others, is required for the localization of viral and cellular proteins to replication compartments (de Bruyn Kops and Knipe, 1988). Viral DNA replication is important for expression of the L genes, and leads to the formation of concatemers of viral genomes. Capsid assembly and DNA packaging also occur in the nucleus. At this stage, capsids are filled with unit length viral genomes, which are produced by cleavage at *pac* cleavage sites in the concatameric DNA (Deiss, Chou, and Frenkel, 1986).

The last stage in virion morphogenesis is the stage of capsid envelopment and virion egress. In the most agreed upon model, the capsid undergoes primary envelopment at the inner nuclear membrane and then loses this first envelope by fusing with the outer nuclear membrane (reviewed in Mettenleiter, Klupp, and Granzow, 2009). This releases the capsid into the cytoplasm where tegumentation occurs and the final envelope is acquired by budding into vesicles derived from the trans-Golgi network (reviewed in Mettenleiter, Klupp, and Granzow, 2006; Turcotte, Letellier, and Lippé, 2005) or endosomes (Das, Vasanji, and Pellett, 2007). Another model is that the capsid acquires its envelope at the inner nuclear membrane where tegument proteins and glycoproteins have accumulated (Epstein, 1962). The particle then travels through the endoplasmic reticulum and acquires a second envelope upon release into the cytoplasm. The particle loses this second envelope by budding through the cell surface. The third model is based on the theory of intracisternal transport (Wild et al., 2002). In this model, capsids are thought to be released from the nucleus through enlarged nuclear pores and are enveloped at the outer nuclear membrane or the rough endoplasmic reticulum. They are then transferred to the Golgi network, and packaged in cytoplasmic vacuoles, which transport the particles to the cell membrane where they are released through exocytosis (Wild et al., 2005). At this time, the envelopment, deenvelopement and reenvelopment is the most widely accepted model.

These steps outline lytic infection, which results in the production of progeny virion that can then infect other neighbouring cells.

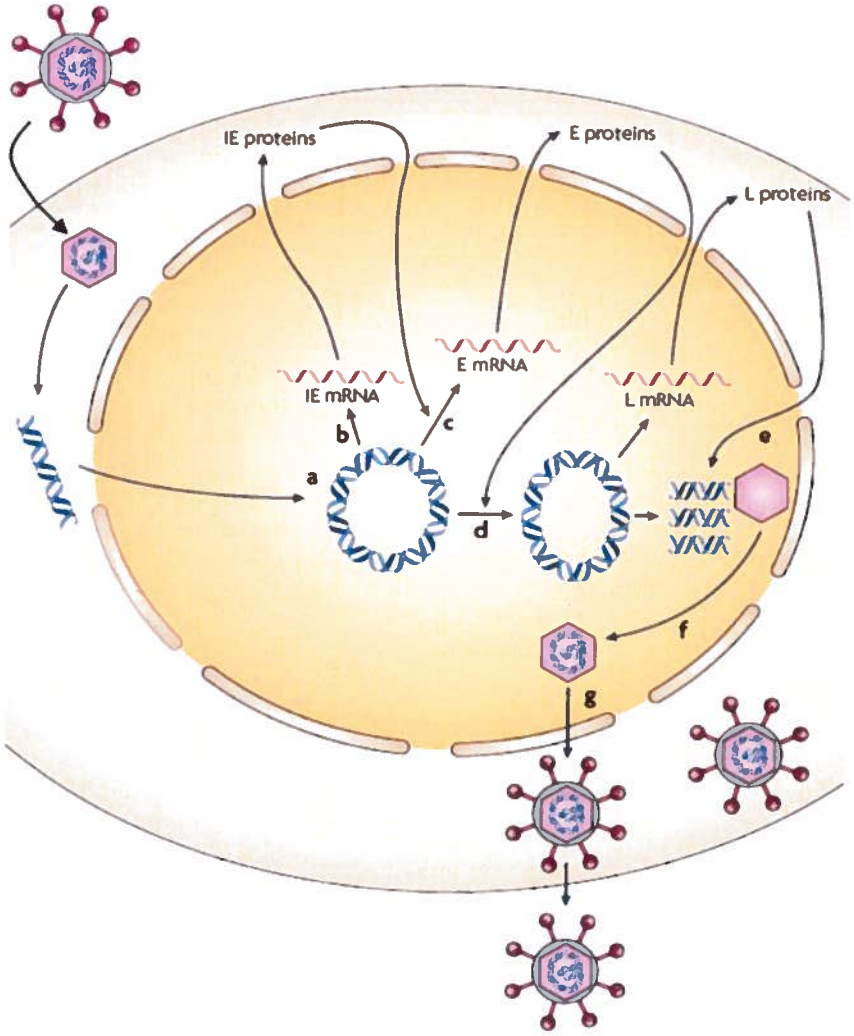

Fig. 3. The viral life cycle. Schematic representation of the viral life cycle: (**a**) viral entry followed by the release of viral genome into the nucleus and circularization of the viral genome, (**b**) transcription and expression of IE genes, (**c**) transcription and expression of E genes, (**d**) viral DNA replication (**e**) transcription and expression of L genes, (**f**) assembly of viral genome into capsids, and (**g**) viral egress. (modified from Knipe and Cliffe, 2008)

1.2 Genetic regulation

1.2.1 Gene expression during HSV-1 infection

HSV-1 codes for over 80 genes, and protein expression often correlates with levels of viral mRNA transcription and accumulation (Jones and Roizman, 1979). In order to understand the tightly regulated cascade of gene expression, many have studied the promoter regions of the different classes of genes, namely IE, E and L. Genes from all three classes can have a TATA element, but IE genes have another element further upstream, the TAATGARAT element. During IE gene activation, the tegument protein VP16 forms a complex with the octamer transcription factor 1 (Oct-1) and the host cell factor (HCF) (Hughes, La Boissiere, and O'Hare, 1999). HCF, binds VP16 and is thought to help in its nuclear import (LaBoissiere and O'Hare, 2000). HCF and VP16 then bind Oct-1, which is already bound to the TAATGARAT element. Together, they stimulate the assembled transcription machinery. In this way, the IE genes coding for ICP4, ICP0, ICP27, ICP22 and ICP47 can be expressed without any *de novo* viral protein synthesis (reviewed in Weir, 2001) using the cellular DNA-dependent RNA pol II. Another important tegument protein is the virion host shut-off (vhs) protein also known as UL41. It degrades pre-existing, as well as, newly generated mRNAs during the first few hours after infection (Kwong, Kruper, and Frenkel, 1988; Strom and Frenkel, 1987). Since the transcription rate of the viral mRNA is higher than that of the cellular mRNA, viral transcripts can accumulate preferentially under this pressure (Taddeo, Esclatine, and Roizman, 2004).

ICP4, ICP0, ICP27 and ICP22 are able to stimulate the transcription of other classes of viral genes (reviewed in Wagner and Bloom, 1997). ICP27 is essential for lytic replication and is involved in the switch from early to late gene expression. It is involved in regulation at the transcriptional and the post-transcriptional level: it plays a role in polyadenylation site selection and 3' mRNA processing (McGregor et al., 1996), inhibition of mRNA splicing (Bryant et al., 2001), RNA binding (Ingram et al., 1996) and in the stimulation of the transport of intronless RNA out of the nucleus and into the cytoplasm (Johnson and Sandri-Goldin, 2009; Phelan, Dunlop, and Clements, 1996). The IE gene ICP47, does not seem to play a role in transcription initiation; however it plays an important role in viral evasion of the immune response by blocking the presentation of peptides at the surface of infected cells (Fruh et al., 1995).

The E proteins that together are necessary and sufficient for viral DNA replication are: UL30 and UL42, which form the DNA polymerase holoenzyme (Gottlieb et al., 1990); ICP8, which is the ssDNA binding protein (Weller et al., 1983); UL9, which is the ORI binding protein and possesses limited helicase activity (Olivo, Nelson, and Challberg, 1988); and UL5/UL8/UL52, which form the heterotrimeric helicase/primase complex (Crute et al., 1989). Viral DNA replication begins

at one of the ORI sites of the genome. The model for DNA replication leading to the production of new viral genomes that is most agreed upon is the rolling-circle mechanism (Jacob, Morse, and Roizman, 1979; Skaliter et al., 1996).

Following the start of DNA replication, E gene expression is either significantly reduced or shut-off while L genes begin to be maximally expressed. Unlike IE and E gene promoters, L gene promoters do not have cis-acting regulatory elements upstream of the TATA element; however, some do contain a downstream activator sequence (Guzowski, Singh, and Wagner, 1994; Guzowski and Wagner, 1993). ICP4 has been reported to play an important role in the activation of E and L genes. Late gene expression is highly activated by ICP4, as late genes contain an initiator element to which ICP4 binds, and stimulates their transcription (Cook et al., 1995; Gu and DeLuca, 1994; Kim, Zabierowski, and DeLuca, 2002). A suggested model that would explain the switch from E to L gene expression is that the transcription complex that forms after DNA replication is different than that formed earlier on during infection, and preferentially binds the elements present in the promoter regions of the late genes (reviewed in Wagner, Guzowski, and Singh, 1995). The preferential expression of late genes at later times in infection is possibly explained by the presence of viral replication compartments where transcription takes place (de Bruyn Kops and Knipe, 1988; Knipe et al., 1987), and which may preferentially contain components required for expression of late genes over the early genes (reviewed in Wagner and Bloom, 1997). A distinction is made between late genes based on the presence or not of low transcription levels before viral DNA replication; leaky-late genes show some transcriptional activity before the start of DNA replication but are maximally expressed after that point, while strict-late transcripts are only detected after the onset of viral DNA replication (reviewed in Weir, 2001). Late gene products are important for encapsidation of viral DNA and for envelopment among other activities.

Genetic regulation is also a determining factor in lytic or latent infections, where it is proposed that the expression of the viral genome is controlled by chromatin structure (Kent et al., 2004). When chromatin is in the euchromatin, less condensed form, genes can be expressed, as is the case during lytic infection; however, during latent infection, the viral DNA is packaged as heterochromatin, a highly condensed form of chromatin, and gene expression is silenced (reviewed in Knipe and Cliffe, 2008). The latency-associated transcript (LAT), highly expressed in latently infected neurons, promotes heterochromatin formation and reduces euchromatin on lytic gene promoters (Wang et al., 2005).

The large and highly dense genome of HSV-1, as well as the necessity to switch between lytic and latent patterns of gene expression, has translated into an elaborate system of gene expression that must be tightly regulated.

1.2.2 Post-transcriptional regulation in eukaryotes

Genetic regulation through regulation of RNA processing can take place at different stages in the path towards protein synthesis: transcription initiation and elongation, splicing, 5' and 3'end processing, RNA stability, nuclear-cytoplasmic export, and finally translation (Ford, Bagga, and Wilusz, 1997; Vinciguerra and Stutz, 2004; Wickens, Anderson, and Jackson, 1997). These processes can be regulated in different ways to control RNA and ultimately protein expression.

Alternative splicing gives rise to transcripts that include different segments of the pre-mRNA depending on which intron(s) are processed and spliced. This mechanism gives rise to mRNAs that differ in their ORF, or their 3' untranslated region (3'UTR), or both (reviewed in Le Hir and Seraphin, 2008). Alternative cleavage and polyadenylation gives rise to transcripts that differ in their 3' ends. Also, the poly(A) tail added to the transcript plays an important role in the export, translation and stability of the mRNA. Deadenylation is the rate limiting step in the decay of most mRNAs, and in higher eukaryotes, it is the major degradation pathway (Mauxion et al., 2009). Efficient capping is also important as the interaction between the 5' cap and the 3' poly(A) tail affects translation initiation (Kuersten and Goodwin, 2003). These are all examples of post-transcriptional genetic regulation.

Regulation of processing at the 3'end of RNA can occur at many levels. *Cis*-acting elements, sequences contained within the 3' UTR of pre-mRNAs, play a very important role in processing efficiency. This is because transcripts generated by RNA pol II are cleaved and polyadenylated by the 3' end processing machinery through recognition of their *cis*-acting elements. Also, the cleavage and polyadenylation processes cannot be separated *in vivo* these two steps are tethered to each other (reviewed in Millevoi and Vagner, 2010). This link is also evident in the function of RNA pol II; along with its role in transcription, RNA pol II interacts with polyadenylation factors and helps deliver them to the 3'end processing complex (Barilla, Lee, and Proudfoot, 2001; Hirose and Manley, 1998). Conversely, components of the 3'end processing complex interact with the RNA pol II to help control transcription initiation, and the polyadenylation signal (PAS) itself plays an important role in transcription termination (Edmonds and Abrams, 1960; McCracken et al., 1997).

The numerous factors implicated and the complexity of the different stages in gene expression provide opportunities for genetic regulation at many different levels.

1.3 The polyadenylation machinery

The polyadenylation machinery comprises the mRNA *cis*-acting elements as well as six polyadenylation factors (Fig. 4). Though the roles of these factors and their subunits have been studied extensively, there is still much to discover.

1.3.1 mRNA *cis*-acting elements

The core elements at the 3'end of transcripts to be processed by the polyadenylation machinery have been defined as: the hexameric AAUAAA consensus sequence, also known as the polyadenylation signal (PAS), the cleavage site at 10-35 nucleotides (nt) downstream from the PAS, and the G/U rich element 14-70 nt downstream of the PAS (reviewed in Lutz, 2008; Takagaki and Manley, 1997). There are also the two auxiliary elements which are more variable but increase signal strength nonetheless: the U-rich element UUUU (Zhao, Hyman, and Moore, 1999) or similar sequences (UGUA, UAUA) (Hu et al., 2005) upstream of the PAS and the G-rich downstream element (DSE) which seems to lack conserved sequence and distance from the PAS (reviewed in Mandel, Bai, and Tong, 2008).

1.3.2 The polyadenylation factors

One of the main factors is the cleavage and polyadenylation specificity factor (**CPSF**). It has five subunits (CPSF-160, CPSF-73, CPSF-100, CPSF-30 and hFip1) all of which are required for efficient cleavage and polyadenylation (reviewed in Mandel, Bai, and Tong, 2008). CPSF-160 is the subunit that binds the hexameric sequence; it is required for polyadenylation and helps direct CPSF-73 to the cleavage site. Efficient binding of CPSF-160 to the transcript depends on the Cleavage and Stimulation Factor (CstF) (see below). CPSF-160 also plays a role in transcription initiation (Dantonel et al., 1997), elongation (McCracken et al., 1997) and termination (reviewed in Proudfoot, 2004). The CPSF-73 subunit is part of the β-CASP family. Members of the β-CASP family are involved in nucleic acid binding and processing, and several members show nuclease activity (reviewed in Dominski, 2007). Therefore, CPSF-73 is most likely the endonuclease for the cleavage reaction. Alone, CPSF-73 has little sequence specificity; its specificity is conferred through the function of CPSF-160 and CstF (reviewed in Mandel, Bai, and Tong, 2008; Mandel et al., 2006). The subunit CPSF-100 is also a member of the β-CASP family but lacks the catalytic site, so it is most likely not an endonuclease (Aravind, 1999; Callebaut et al., 2002). Its function is not yet determined but it has been proposed to act as a scaffolding protein (Kyburz et al., 2003). The CPSF-30 subunit preferentially binds a poly(U)-rich sequence *in vitro* (Barabino et al., 1997), and was shown to bind the U-rich upstream auxiliary element in

yeast (Barabino, Ohnacker, and Keller, 2000). The remaining subunit is hFip1, and its primary function may be to bring the polyA polymerase (PAP) close to the PAS.

CstF is important for cleavage and polyadenylation, and contains three subunits, CstF-50, CstF-64 and CstF77. The CstF-64 subunit has an RNA-binding domain, and was shown to preferentially bind the G/U-rich DSE (reviewed in Takagaki and Manley, 1997). It appears to be a common target for the regulation of gene expression (reviewed in Danckwardt, Hentze, and Kulozik, 2008). One example is the switch of IgM heavy-chain protein expression from the membrane bound to the secreted form. This switch is caused by an increase in CstF-64 expression likely leading to favoured processing of the weaker promoter proximal PAS within the ORF and thus the expression of the shorter, secreted IgM (Takagaki et al., 1996). It has also been shown that there is a 5-fold increase in CstF-64 expression during the transition from the G_0 to S phase in the cell cycle in 3T6 fibroblasts and in primary human splenic B cells (Martincic et al., 1998). The CstF-77 subunit is required for proper 3'end cleavage and may function as a dimer (reviewed in Mandel, Bai, and Tong, 2008). CstF-77 and CstF-50 have been shown to bind the RNA pol II C-terminal domain (CTD) (McCracken et al., 1997). CstF-50 has also been shown to be required for cleavage *in vitro* (Takagaki and Manley, 1994).

The mammalian cleavage factor I (**CFI_m**) is important for cleavage and functions as a heterodimer with four subunits: the 25, 59, 68-kDa polypeptides, as well as the less abundant 72-kDa subunit. These subunits copurify with CFI_m activity from HeLa cell nuclear extracts (Ruegsegger, Beyer, and Keller, 1996). The 25-kDa subunit interacts with the PAP (Kim and Lee, 2001) and the polyA binding protein (PABP) (Dettwiler et al., 2004). Interestingly, despite the presence of an RNA recognition motif (RRM) on the 68-kDa subunit, this subunit does not bind RNA very strongly, rather, it is the domain that is necessary for binding to the 25-kDa subunit, which does bind RNA (Kim and Lee, 2001).The CFI_m has also been identified as a component of purified spliceosomes and the 68-kDa subunit has a domain organization similar to spliceosomal SR proteins containing an RRM (Fribourg et al., 2003).

The mammalian cleavage factor II (**$CFII_m$**) also important for cleavage has two subunits: hClp1 (47 kDa) and hPcf11 (140-200 kDa). hClp1 appears to play a role in bridging the two factors CFI_m and CPSF of the cleavage and polyadenylation machinery. Antibodies directed against hClp1 abolish cleavage, but not polyadenylation activity from HeLa cell nuclear extracts suggesting a role for hClp1 in cleavage (de Vries et al., 2000). The hPcf11 N-terminal portion interacts with the CTD of RNA pol II and preferentially binds the phosphorylated form of the enzyme (Barilla, Lee, and Proudfoot, 2001; Zhang and Gilmour, 2006) – an interaction that seems to play an important role in transcription termination (Hollingworth et al., 2006). The C-terminal portion of hPcf11 has not yet been characterized (reviewed in Mandel, Bai, and Tong, 2008).

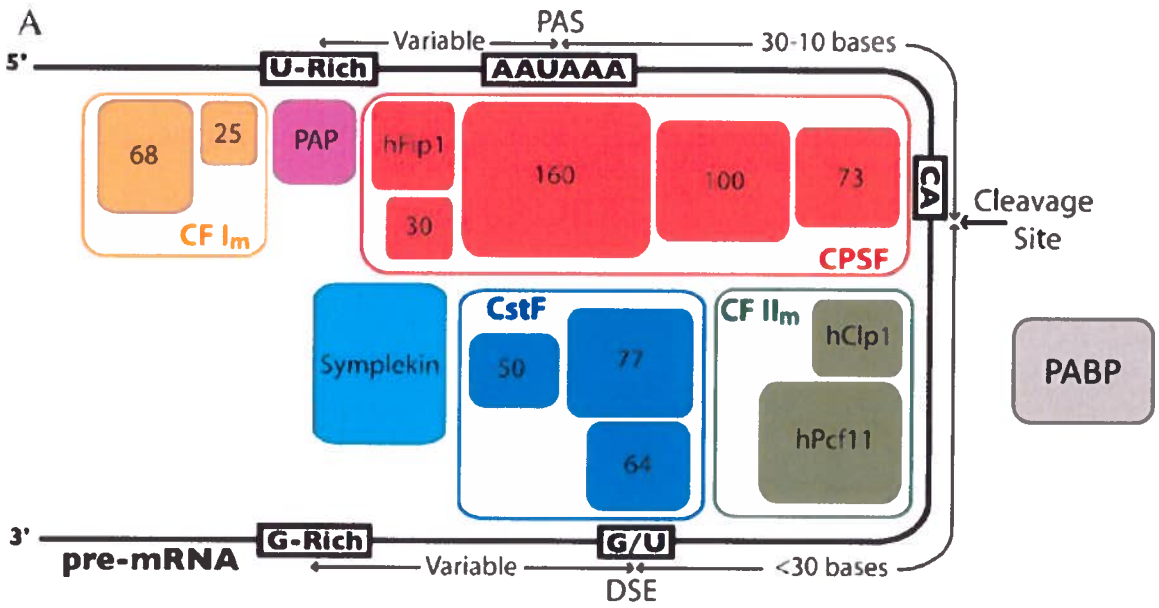

Fig. 4. The polyadenylation machinery. Diagram showing an overview of the polyadenylation machinery. The black line depicts the 3' end of pre-mRNA and the different *cis*-acting elements. The various polyadenylation factors and their subunits are also shown: CPSF (red box), CstF (blue box), CFI_m (yellow box) and $CFII_m$ (green box), the PAP (purple shape), the PABP (grey) and symplekin (light blue shape). The numbers in the rectangles refer to the subunit names and also correspond to their molecular masses in kilodaltons. (modified from Mandel, Bai, and Tong, 2008)

The polymerase **PAP** is responsible for the addition of the poly(A) tail to pre-mRNAs after cleavage. PAP is an induced-fit enzyme whose conformation changes slightly after substrate binding, which activates the enzyme - this property is important in defining substrate specificity (Balbo and Bohm, 2007; Balbo, Meinke, and Bohm, 2005; Balbo, Toth, and Bohm, 2007; Mandel and Tong, 2007). *In vitro*, the PAP alone shows polyadenylation activity; however, interaction with the other components of the complex is important for defining the poly(A) length (reviewed in Edmonds, 2002).

There are two types of **PABP**s. PABPN1 binds the mRNAs in the nucleus. In the cytoplasm, PABPC replaces PABPN1 on the RNA, where it stimulates translation and regulates mRNA stability (Kahvejian, Roy, and Sonenberg, 2001; Kuhn and Wahle, 2004; Sachs, Sarnow, and Hentze, 1997). PABPN1 has four RRMs (Deo et al., 1999) and is required for correct and efficient polyadenylation. It controls the length of the poly(A) tail by binding to the tail as it forms

(Meyer, Urbanke, and Wahle, 2002) and until it reaches the appropriate length of 200 to 300 bases (Keller et al., 2000). The PABP protein also increases the efficiency of the PAP 80-fold (Kerwitz et al., 2003).

Symplekin appears to function as a scaffolding protein bringing together different components of the cleavage and polyadenylation machinery. It interacts with CstF-64 and forms a stable complex with CPSF and CstF (Takagaki and Manley, 2000).

Together, the different components of the cellular cleavage and polyadenylation machinery form an efficient and versatile system for the processing and maturation of pre-mRNAs.

1.3.3 Mechanisms for regulation of 3'end processing

Several mechanisms for regulating 3'end formation have been identified and compiled (reviewed in Millevoi and Vagner, 2010).

One mechanism is the competition of a regulatory factor with core polyadenylation factors for binding to the PAS (Fig. 5a). Similarly, factors involved in transcription, splicing and polyadenylation can also compete for binding, where binding depends on the kinetics of assembly of the various competing complexes; the process able to assemble its components first will be favoured (Fig. 5c). The alternative scenario is also observed where factors involved in one process help recruit the factors of another process favouring the latter (Fig. 5b). Also, in the presence of suboptimal *cis*-acting elements, the differential expression of polyadenylation factors can redirect the polyadenylation machinery to an alternative PAS (Fig. 5d). Under these conditions, increased expression of polyadenylation factors will lead to better processing of a weaker signal and vice versa. These four methods all involve a change in the association of the polyadenylation factors caused by competition, recruitment or changes in expression levels. It is also possible to have a regulatory factor bind a *cis*-acting RNA element, interact with one of the polyadenylation factors, and inhibit its activity, without affecting the level of association the factor at hand (Fig. 5e).

Another mechanism is one that leads to the redistribution of the polyadenylation factors from the nucleus to the cytoplasm resulting in negative regulation of cleavage and polyadenylation, or to nuclear protein complexes resulting in either negative or positive regulation (Fig. 5f and g).

Finally, regulation can occur through post-translational modifications of both the pre-mRNA 3'end processing factors and regulators leading to either positive or negative regulation (Fig. 5h); a process associated with many of the mechanisms described above. These modifications include methylation, sumoylation, acetylation and phosphorylation, and can modulate the activity, the nuclear/cytoplasmic partitioning, stability and association of the different factors.

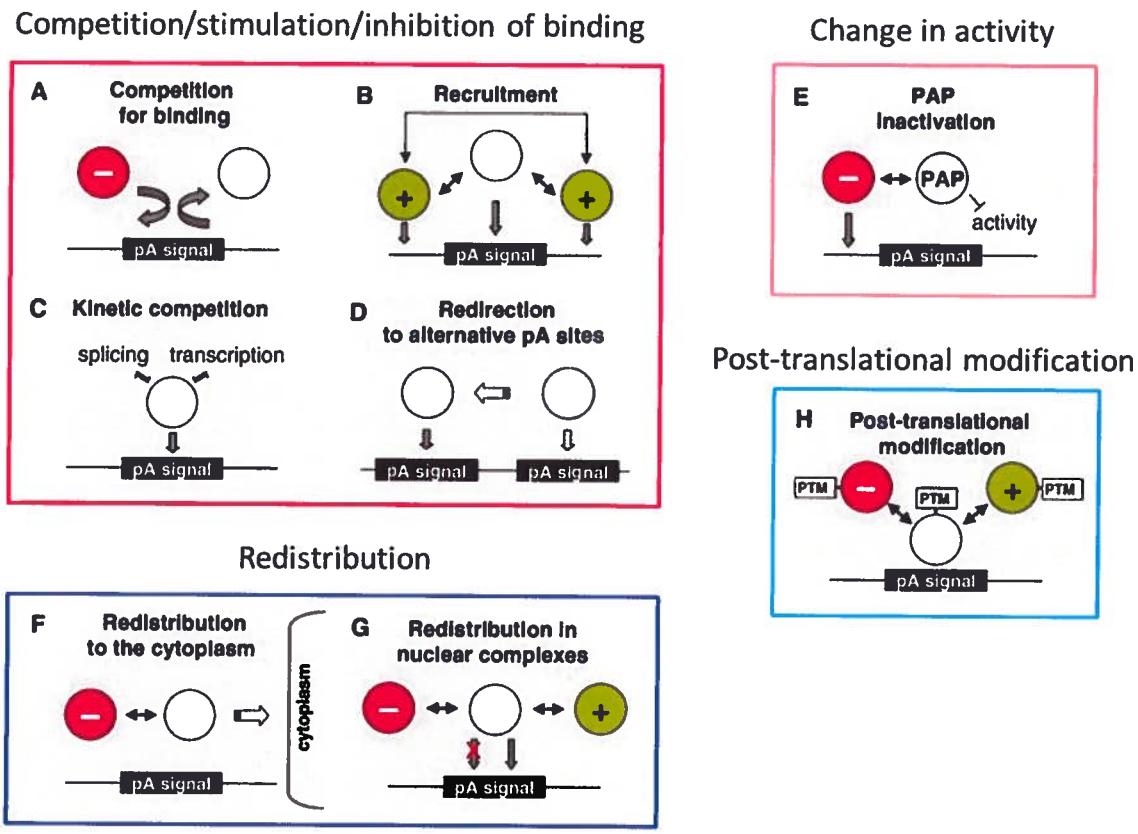

Fig. 5. Mechanisms of post-transcriptional regulation (modified from Millevoi and Vagner, 2010). Diagrams showing the different mechanisms for post-transcriptional regulation categorized. Competition/stimulation/inhibition of binding (red box): (a) competition for binding, (b) recruitment, (c) kinetic competition, (d) redirection to alternative polyadenylation sites. Change in activity (pink box): (e) inactivation of, for example, the PAP. Redistribution (purple box): (f) redistribution to the cytoplasm, (g) redistribution in nuclear complexes. Post-translational modification (blue box): (h) post-translational modification leading to increased or decreased processing.

1.4 Alternative polyadenylation

Alternative polyadenylation is defined as the process by which transcripts generated from a single gene can be cleaved and polyadenylated in different ways due to the presence of more than one PAS, leading to multiple mRNA transcripts (reviewed in Lutz, 2008). If one of the potential polyA sites is within the open reading frame, the length of the protein generated from this process will be affected. If alternative polyadenylation occurs within the 3' UTR, then the stability, localization, transport and translation efficiency of the transcript can be affected (Chao et al., 1999) (Fig. 6). This phenomenon has led to the study of the *cis*-acting elements contained in alternatively polyadenylated signals to help understand components that increase or decrease the strength of a polyadenylation signal.

The combined effect of various post-transcriptional regulation mechanisms including the activities of polyadenylation regulation factors or other regulatory factors, as well as *cis*-acting RNA elements, affect the efficiency with which a polyadenylation site will be processed. Studies predict that alternative polyadenylation is a process that occurs for over half of the human genes (Beaudoing et al., 2000; Tian et al., 2005). Tissue-specific alternative polyadenylation has also been reported and can be triggered by tissue specific expression levels of core polyadenylation proteins (Dass et al., 2001; Dass et al., 2007; Wallace et al., 2004) or even by other factors (Costessi et al., 2006; Winter et al., 2007). These results highlight the importance of alternative polyadenylation in eukaryotic gene expression and in the generation of transcript diversity, and demonstrate the importance of developing a better understanding of this cellular phenomenon.

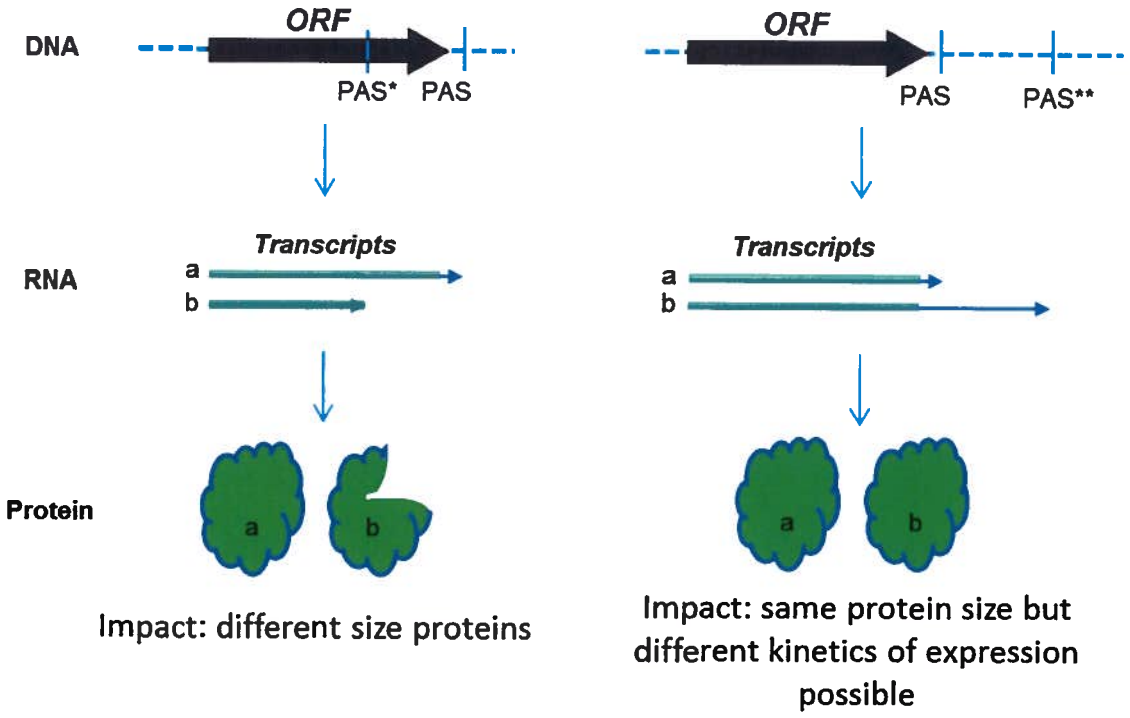

Fig. 6. Alternative polyadenylation. The left panel shows alternative polyadenylation where one of the PAS is within the ORF (black arrow). "PAS" indicates the polyadenylation site that is present right after the ORF, and "PAS*" indicates the polyadenylation site within the ORF. The right panel shows alternative polyadenylation within the 3'UTR, where "PAS**" indicates the polyadenylation site used downstream of the first PAS of the gene. The transcripts generated from these two processes are shown by the blue arrows below with light green rectangles highlighting the ORFs. Proteins sizes are represented by the green clouds in the lower panels.

1.4.1 Alternative polyadenylation during HSV-1 infection

Alternative polyadenylation is used for temporal expression of genes in many viral systems including cytomegalovirus (Goins and Stinski, 1986; Stamminger, Puchtler, and Fleckenstein, 1991), adenovirus (Mann, Weiss, and Nevins, 1993; Nevins and Wilson, 1981), polyomavirus (Hyde-DeRuyscher and Carmichael, 1990) and papillomavirus (Kennedy, Haddow, and Clements, 1990).

During HSV-1 infection, splicing factors are withdrawn from active transcription sites, thus favouring the polyadenylation of transcripts over splicing in cases where the two processes were in competition (Martin et al., 1987). And though not directly related to alternative polyadenylation, it was also shown that certain nuclear proteins bind with more efficiency to viral mRNAs than to cellular mRNAs during HSV-1 infection (McGregor et al., 1996).

ICP27 stimulates 3'end processing at certain polyadenylation sites; it was shown that the usage of the polyadenylation sites of two late genes, *UL38* and *UL44*, was stimulated in the presence of ICP27 while it had no effect on the processing of four other IE and E genes (McGregor et al., 1996). The polyadenylation sites of *UL38* and *UL44* are inherently weak, so it was proposed that ICP27 might favour the use of weak PAS late during the infection; however, an ICP27-null mutation did not have an effect on the expression of *UL24* short transcripts, which use a weak PAS, but rather affected the expression of the long transcripts, which use a relatively strong polyadenylation signal (Hann et al., 1998).

Alternative polyadenylation in HSV-1 has been observed for several genes, namely, *UL2* (Singh and Wagner, 1993), *UL24* (Cook and Coen, 1996), *UL38* (Anderson et al., 1981) and *UL52* (Uprichard and Knipe, 1996).

1.4.2 HSV-1 *UL24*

The *UL24* gene is evolutionarily conserved among many herpes viruses (reviewed in Davison, 2002; Jacobson, Martin, and Coen, 1989). A *UL24*-null HSV-1 strain exhibits decreased viral yield *in vitro* (Jacobson, Martin, and Coen, 1989), and displays a syncytial plaque phenotype (Sanders, Wilkie, and Davison, 1982). In a mouse model of eye infection, a *UL24*-null virus exhibits defects in viral replication and reactivation from latency (Jacobson et al., 1998). Also, the UL24 mutant virus strain vUL24-E99A/K101A, harbouring mutations targeting the UL24 endonuclease motif, produced lower viral titers, fewer clinical symptoms and showed defects in reactivation from latency in a mouse model of eye infection (Leiva-Torres, Rochette, and Pearson, 2010).

During HSV-1 infection, viral replication compartments are formed in the nucleus through remodelling of the nucleus by viral proteins (Maul, Guldner, and Spivack, 1993; Monier et al., 2000; Reynolds, Liang, and Baines, 2004). UL24 plays an important role in HSV-1-induced nucleolar modifications. It was shown to be involved in dispersal of the nucleolar proteins, nucleolin and B23, in the context of transfection as well as infection (Bertrand et al., 2010; Bertrand and Pearson, 2008; Lymberopoulos and Pearson, 2007).

UL24 generates six types of transcripts (Cook and Coen, 1996) (Fig. 7). They are transcribed from three different start sites, one of which is within the *UL24* ORF (Kibler et al., 1991; McGeoch et al., 1988; Read, Sharp, and Summers, 1984; Read and Summers, 1982; Wilkie et al., 1980). Transcripts from these three different start sites utilize either the *UL24* PAS to generate a set of short transcripts or the *UL26* PAS to generate a set of long transcripts (McGeoch et al., 1988). The short transcripts are expressed with early kinetics while the long transcripts are expressed with leaky-late kinetics (Cook and Coen, 1996). Regulation of *UL24* transcript expression is complex and is affected by antisense regulation by the HSV-1 thymidine kinase (*tk*) mRNA, which overlaps the *UL24* ORF (Cook et al., 1996).

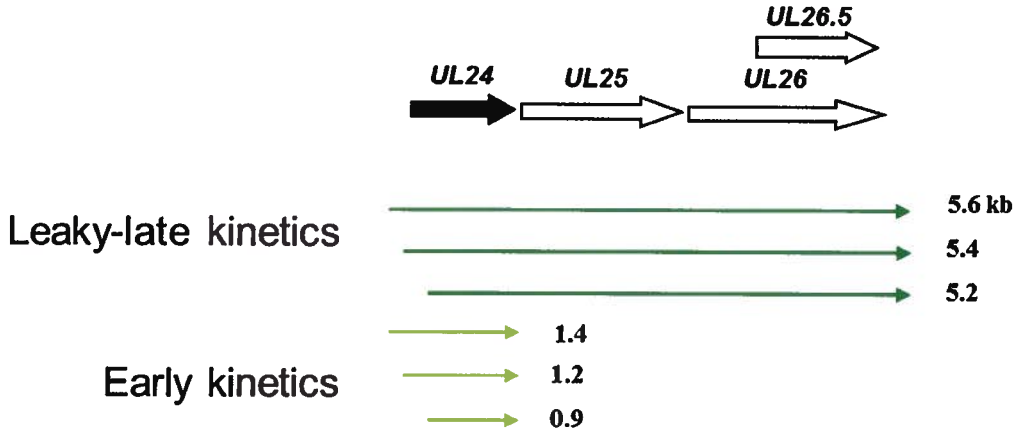

Fig. 7. The *UL24* gene and its transcripts. The thick black arrow represents the *UL24* ORF. The white arrows represent the ORFs of *UL25*, *UL26* and *UL26.5*. The thin dark green arrows represent the long transcripts and the light green arrows represent the short transcripts. (Hann et al., 1998)

The long transcripts also contain the coding information for *UL25*, *UL26*, *UL26.5*. *UL25* is important for the packaging of viral DNA (McNab et al., 1998; Newcomb, Homa, and Brown, 2006). *UL26* codes for a polypeptide that is cleaved to generate the viral capsid protein and protease VP24, as well as, the scaffolding protein 21 (Liu and Roizman, 1991; Preston et al., 1992). *UL26.5* codes for the capsid protein 22a (Kennard et al., 1995; Liu and Roizman, 1991).

The UL24 protein is expressed with leaky-late kinetics (Pearson and Coen, 2002). A mutation that eliminated the expression of the transcripts starting at the first transcription start site resulted in a 10-fold decrease, but not complete elimination, of UL24 expression (Pearson and Coen, 2002). Moreover, infection with an ICP27-null virus mutant resulted in undetectable amounts of UL24 protein (Pearson, Knipe, and Coen, 2004). This mutant virus also showed reduced amounts of the *UL24* long transcripts but had little effect on the level of short transcripts (Hann et al., 1998). Also, the effect of ICP27 on transcript localization was studied. With the wild-type virus, short as well as long transcripts localized primarily in the cytoplasm (Pearson, Knipe, and Coen, 2004). However, with the ICP27-null virus, more than half of the long transcripts remained in the nucleus, while the short transcripts seemed to be exported with the same efficiency (Pearson, Knipe, and Coen, 2004). This result also suggests that the *UL24* short transcripts are poorly translated because although they were formed in the cytoplasm, UL24 protein expression was undetectable (Pearson, Knipe, and Coen, 2004). Therefore, UL24 expression seems to correlate with the expression of the long transcripts.

The function of the short and long transcripts generated, as well as how their expression is regulated, remains unclear.

1.4.3 HSV-1 *UL38*

UL38 expression is also regulated by alternative polyadenylation. UL38 is essential for capsid assembly *in vitro* (Tatman et al., 1994; Thomsen, Roof, and Homa, 1994). It codes for VP19C, which forms a complex with VP23 (Newcomb et al., 1993; Trus et al., 1996; Zhou et al., 2000), and interacts with the major capsid protein VP5 (Desai and Person, 1996). *UL38* transcripts are generated through use of the *UL38* PAS, as well as, a polyadenylation signal further downstream, that shared by *UL39* and *UL40* (Fig. 8). UL39 and UL40 are the two components of the viral ribonucleotide reductase: RR1 and RR2 respectively (Anderson et al., 1981). The *UL38* transcription pattern has not been studied as extensively as that of *UL24* however, it has been shown that ICP27 increases the 3'end processing of the inherently weak UL38 PAS (McGregor et al., 1996).

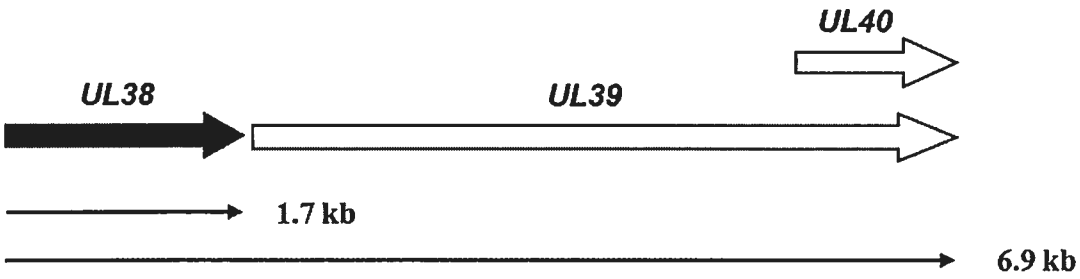

Fig. 8. The *UL38* gene and its transcripts. The thick black arrow represents the ORF of *UL38*. The black and white arrows represent the ORFs of *UL39* and *UL40*. The thin solid lines represent the short and long transcripts of *UL38*. (Anderson et al., 1981)

1.5 Purification techniques for RNA-protein complexes

Protein-bound RNAs are present in the nucleus as well as the cytoplasm, and the proteins they bind influence their processing, transport, localization, translation and stability (reviewed in Dreyfuss, Kim, and Kataoka, 2002).

Advances in techniques for protein purification and isolation of protein complexes to help determine protein function and chemistry has led to the development of protein affinity tags like the polyhistidine tag (Porath, 1992; Porath et al., 1975), the FLAG aptamer (Prickett, Amberg, and Hopp, 1989) and glutathione S-transferase (GST) (Simons and Vander Jagt, 1981) etc. These tags bind with high affinity to a ligand that can be immobilized on a chromatography resin. The isolated complex can be eluted either by a competing ligand or under denaturing conditions. In some cases, the tag can be cleaved by a protease to release the bound complex. This well established concept has since been extrapolated to the purification of complexes with nucleic acids including RNA.

There are several methods used to tag RNA for subsequent analysis (Walker et al., 2008): (1) chemical tagging during *in vitro* transcription, (2) incorporation of a well characterized RNA protein-binding sequence, (3) hybridization of an affinity-tagged oligonucleotide, (4) the 3-hybrid system, and (5) incorporation of an *in vitro* selected RNA protein-binding sequence.

1.5.1 Chemical Tagging

Chemical tagging of *in vitro* transcribed RNA can be done through incorporation of modified ribonucleoside triphosphates. These chemical modifications include the addition of biotin (Rouault et al., 1989), fluorescent dyes (Proudnikov and Mirzabekov, 1996; Roget, Bazin, and Teoule, 1989) and digoxeginin (Holtke et al., 1995). The two main drawbacks of this technique are that these chemical modifications can impact RNA-protein interactions by causing changes in RNA structure that inhibit complex formation, and also, this type of RNA tagging is limited to *in vitro* studies (Walker et al., 2008).

1.5.2 Characterized RNA-binding protein sequence tag

There are several well characterized RNA-protein interactions that have been exploited in the development of RNA tags, namely, the bacteriophage MS2 coat protein (Bardwell and Wickens, 1990), the U1 small nuclear ribonucleoprotein A (Liang and Lutz, 2006), as well as, the pseudomonas phage PP7 coat protein (Hogg and Collins, 2007), and their respective cognate RNAs. Ironically, the main challenge in the use of these types of tags for affinity purification is the elution efficiency, which is compromised by the typically strong RNA-protein affinity. One way to circumvent this issue is to fuse a secondary protein like the maltose-binding peptide, to the primary protein and elute using maltose as the ligand (Das, Zhou, and Reed, 2000). The alternative is to engineer a protease cleavage site between the peptide and the protein, which can be used to release the complex for elution (Leonov et al., 2003). This technique typically requires the elution to be performed under denaturing conditions.

1.5.3 Biotinylated oligonucleotide tags

A biotinylated oligonucleotide sequence that is complementary to an accessible segment of the RNA to be evaluated can also be used to purify RNA-protein complexes (Blencowe et al., 1989; Lingner and Cech, 1996). Elution using this technique can be achieved through addition of a competitive oligonucleotide sequence or through denaturation. This technique avoids possible misfolding of the RNA caused by chemical modifications or by the addition of a foreign sequence, and furthermore, no genetic engineering is required. While the accessibility of the different sites can be quickly assessed, an important limiting factor for the use of this technique is the complexity of either the RNA or of the ribonuclearprotein, since an unstructured and accessible region is essential in allowing for interaction with the oligonucleotide sequence (Walker et al., 2008).

1.5.4 3-Hybrid system

In the 3-hybrid system (Bernstein et al., 2002), a hybrid-RNA sequence is designed to be able to bind two proteins with high affinity (Fig. 9). Both proteins are in fact hybrid proteins with two components, one that binds the hybrid RNA, and a second protein with either the DNA binding domain or the activation domain of a transcriptional activator. If all components of the system are present and functional, the promoter will be activated and the reporter gene will be transcribed and expressed.

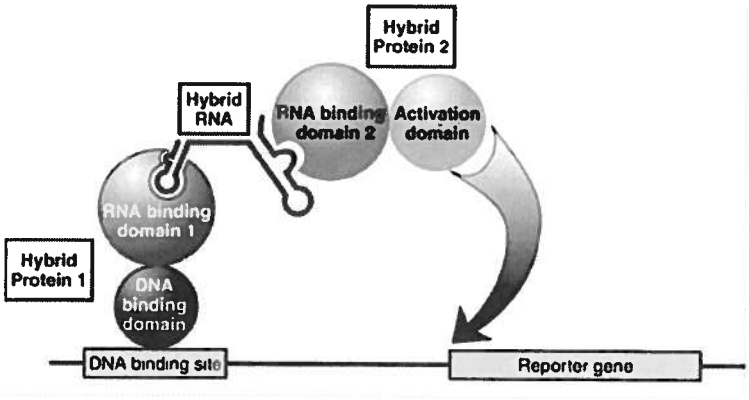

Fig. 9. The 3-Hybrid system. Diagram showing the components of the 3-hybrid system: the hybrid RNA, the hybrid protein 1 and 2 and the reporter gene. (modified from Bernstein et al., 2002)

1.5.5 Artificial RNA affinity tags

In vitro selection has led to the discovery of RNA or DNA sequences, named aptamers, which can efficiently bind small molecules or macromolecules (Wilson and Szostak, 1999). One possible application of these aptamers is in affinity purification of RNA-binding proteins (Walker et al., 2008). Two systems have been detailed by Srisawat and Engelke, who have discovered high affinity tags that bind streptavidin or Sephadex (Srisawat and Engelke, 2001; Srisawat and Engelke, 2002; Srisawat, Goldstein, and Engelke, 2001). The resins are commercially available in each case; however, the Sephadex beads are a relatively economical choice, whereas the streptavidin resin is more expensive. Also, being that the concentration of ligands on the beads is high in the Sephadex system, it may be more suitable for large scale purification than the streptavidin system. Both systems offer the option of specific elution, as well as elution under denaturing conditions; however, in the Sephadex system, removing the dextran used to elute under non-denaturing conditions is more difficult than removing the biotin counterpart in the streptavidin system. Additionally, the streptavidin aptamer distinguishes itself from the Sephadex aptamer by its high affinity (70nM) and low background (Fig. 10). Although biotin, streptavidin's natural ligand is endogenously present in cell lysates, it can be blocked by the addition of egg white avidin before loading the lysates on the column. As is the case for many of the techniques listed here, the underlying RNA-protein interaction is stable and strong under specific conditions. Thus changes to these conditions due to experimental constraints of the samples to be analysed must be tested and optimized. Also, special care must be taken to ensure the RNA tags are not degraded by RNases present in samples, buffers or any material used. Despite some limitations, these aptamers allow for rapid, specific and efficient purification of specific RNA-binding proteins.

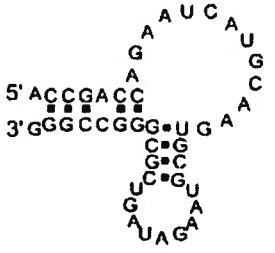

Fig. 10. The streptavidin aptamer. Nucleotide sequence and 2-D structure of the streptavidin S1 aptamer.

CHAPTER 2: RESEARCH QUESTION

HSV has a double stranded DNA genome that codes for more than 80 genes, necessitating effective genetic regulation. One of the modes of regulation utilized by HSV-1 is alternative polyadenylation: a post-transcriptional event that allows one gene to use two different PAS. This process generates transcripts that differ in their 3'UTR, which can affect translation efficiency among other things. One of the HSV-1 genes whose expression is regulated by alternative polyadenylation is *UL24*. Short *UL24* transcripts arise from use of the *UL24* polyadenylation signal and are expressed with early kinetics. Long, polycistronic *UL24* transcripts arise from use of the *UL26* polyadenylation signal, and these transcripts exhibit late kinetics. Similarly, the *UL38* gene also has two types of transcripts that arise from alternative polyadenylation, the short transcripts use the *UL38* polyadenylation signal and the longs transcripts use the *UL39/UL40* polyadenylation signal further downstream.

We hypothesized that (I) the switch from production of short to long transcripts is due to changes in the steady-levels of polyadenylation factors over the course of infection. We also hypothesized that (II) alternative polyadenylation was mediated in part by the degree of association of regulatory RNA-binding proteins with the different polyadenylation sites, and that these interactions change over the course of infection.

To test these hypotheses, we analysed lysates from Vero and HeLa cells either mock-infected, or infected for 3, 6, 9, 12 and 16 hrs. Levels of various cellular polyadenylation factors were assessed by Western blot. In addition, an RNA affinity approach was developed to identify proteins that interact with the 3'UTRs of the long and short transcripts.

CHAPTER 3: MATERIAL AND METHODS

3.1 Cells and virus

The HSV-1 strain KOS (Davar et al., 1994) was originally obtained from Donald M. Coen (Harvard Medical School). Vero cells (African green monkey kidney cells) were grown in Dulbecco's Modified Eagle's Media (DMEM) with 5% newborn calf serum (NCS), and penicillin and streptomycin (Pen/Strep). HeLa cells (human epithelial cervical cancer cells) were grown in DMEM with 8% fetal bovine serum (FBS), and penicillin and streptomycin. Cells were maintained in a humidified incubator with 5% CO_2 at 37 °C.

3.2 Antibodies

Most primary antibodies used were obtained from Abcam: CPSF-73 (ab72294), CPSF-100 (ab81553), CPSF-30 (ab51343), CstF-2 (ab72297), CstF-3 (72299), CPSF5 (ab66667), CPSF6 (ab56858). CPSF7 (sc-133481) was obtained from Santa Cruz. The antibody directed against CPSF-160 (BP301-580) was obtained from Bethyl Laboratories. The anti-PABP antibody was obtained from Millipore (05-847).

The goat anti-mouse and goat anti-rabbit secondary antibodies conjugated to horseradish peroxidase were purchased from SIGMA (A4416) and Bethyl Laboratories (A120-101P) respectively.

3.3 Western blots

The day before infection, $2x10^6$ Vero cells or $3.5x10^6$ HeLa cells were seeded in 60 mm dishes. The next day, cells were either mock-infected or infected with KOS virus at an MOI of 5. The inoculum was prepared in DMEM with 2% NCS and Pen/Strep. To infect the cells, the growth medium was removed and 4 ml of warm (37°C) inoculum were added to each dish. The dishes were incubated at 37°C for 1 hr with shaking every 15 min. The cells were then washed twice with warm PBS, and then PBS was replaced with growth medium. The cells were incubated for 3, 6, 9, 12 or 16 hpi. The cells were then placed on ice and washed twice with cold PBS. To each dish, 500 ul of RIPA lysis buffer (50 mM Tris-HCl pH 7.5, 1% Triton X-100, 0.5% deoxycholate, 0.1% SDS), with 500 mM NaCl were then added, which were incubated on ice with occasional shaking for 15 min. The lysed cells were then scrapped and centrifuged at top speed in a microcentrifuge for 30 min at 4°C to remove cellular debris. The supernatants were transferred to a fresh tube and stored at -80°C.

Lysates were resolved on a 10% polyacrylamide SDS gel and proteins were transferred to a polyvinylidene fluoride membrane using the Bio-Rad Trans-Blot Electrophoretic Transfer Cell. Transfer was carried out for 1.5 hrs in transfer buffer (25 mM Tris base, 192 mM Glycine and 20% MeOH). Membranes were blocked for 1 hr in a blocking solution of 5% nonfat dry milk (Blotting-Grade Blocker, BioRad) prepared in TTBS (50 mM Tris-HCl, pH 7.4, 150 mM NaCl and 0.5% Tween). Primary antibody solutions were diluted 1:2000 in TTBS with 0.5% gelatin (Fisher). Secondary antibody solutions were diluted 1:5000 in blocking solution. Detection was by Enhanced chemiluminescence plus (GE) according to the manufacturer's instructions.

3.4 Plasmids

Templates for in vitro transcription were based on the vector pBluescript SK+ (Stratagene). The polyadenylation signals of HSV-1 strain KOS *UL24*, *UL26*, *UL38* and *UL40* were each inserted upstream and in tandem with DNA coding for the S1-tag. The orientation of the fragments was such that their transcription was under the control of the T3 bacteriophage promoter. pSK-pAUL24 and pSK-pAUL26 contain the *UL24* PAS, an approximately 140 bp sized fragment, and the *UL26* PAS, an approximately 220 bp sized fragment, respectively, subcloned into the PstI site of pBluescript SK+ (A. Pearson, unpublished data).

3.4.1 pSK-pAUL38 and pSK-pAUL40 plasmid construction

The same general strategy was used for the construction of the plasmids containing the segments of the 3'UTRs containing the polyadenylation regions of *UL38* and *UL40*. A segment of the 3' UTR containing the polyadenylation region of *UL38* was amplified from KOS viral DNA by PCR using the primers 5' CTG CAG TGA GCG TAG CAA ACG CC 3' and 5' CTG CAG GCA TTT TGT GGC TTC CTC 3' (Integrated DNA Technologies). The 3'UTR containing the polyadenylation region of *UL40* was amplified using 5' CTG CAG AGT GTC GCA GCA CCT C 3' and 5' CTG CAG CGT CCC TGA CAA GAA TC 3' as primers. The underlined nucleotides indicate the PstI restriction sites, which were included in the primers to insert PstI restriction sites at each end of the amplified fragments.

pfu DNA polymerase (Fermentas) was used for PCR amplification of this sequence thus generating fragments with blunt ends. The segment amplified for *UL38* was: 5' CTG CAG TGA GCG TAG CAA ACG CCC CGC CCA CAC AAC GCT CCG CCC CCA ACC CCT TCC CCG CTG TCA CTC GTT GTT CGT TGA CCC GGG CGT CCG CCA AAT AAA GCC ACT GAA ACC CGA AAC GCG AGT GTT GTA ACG TCC TTT GGG CGG GAG GAA GCC ACA AAA TGC CTG CAG 3'. The segment amplified for *UL40* was: 5' CTG CAG AGT GTC GCA GCA CCT CCT ACG CCG GGG CGG TCG TCA ACG ATC TGT GAG ***G***GT CGC GGC GCG CTT CTA CCC GTG TTT GCC CAT AAT AAA CCT CTG AAC CAA ACT

TTG GGT CTC ATT GTG ATT CTT GTC AGG GAC GCT GCA G 3'. The guanine, G in bold, in the KOS strain sequence is different than the nucleotide in the strain 17 sequence, which has a thymine at that position instead.

To generate the plasmid containing the 3'UTRs of *UL38* and *UL40*, the pSK+ vector was digested with the SmaI restriction enzyme, a blunt end cutter. The PCR product and the digested vector were then gel purified using the SIGMA GenElute™ gel extraction kit following the manufacturer's instructions. Prior to ligation, the PCR fragment obtained was phosphorylated using the T4 polynucleotide kinase (NEB), and the digested vector was dephosphorylated with the Antarctic phosphatase (NEB) to reduce the background clones due to religated vector.

The PCR products with the 3'UTRs of *UL38* or *UL40* were inserted non-directionally into the SmaI site using T4 DNA ligase (NEB) to generate the plasmids pSK-pAUL38 and pSK-pAUL40 (with inserts in the same orientation as the T3 promoter), as well as, pSK-pAUL38inv and pSK-pAUL40inv (inserts in the opposite orientation of the T3 promoter).

All ligations were carried out using T4 DNA ligase (NEB) and reaction mixtures were incubated overnight at 15-20°C. Ligation reaction products were used to transform DH5α competent cells. Transformation mixtures were plated onto LB agar plates with 50 μg/ml of ampicillin and incubated at 37°C overnight. Ampicilin resistant colonies were used to inoculate 2 ml of LB broth with 50 μg/ml of ampicillin. The plasmids were purified using the GeneJET™ Plasmid Miniprep Kit (Fermentas) according to manufacturer's instructions. The plasmids obtained were screened for the desired insert by digestion with PstI and the digestion products were resolved on 1.2% agarose gels. The DNA samples from positive clones were sent for sequencing (Génome Québec) to ensure no errors were incorporated into the sequence and to determine the orientation of the inserts.

3.4.2 Addition of the S1 aptamer to the plasmid constructs

Next, the constructs containing the S1 aptamer (Srisawat and Engelke, 2001) as well as the PAS were prepared. The insert containing the S1 aptamer sequence was generated by annealing two complementary synthetic oligonucleotides, each with non-complementary ends corresponding to XhoI overhangs. The top sequence 5' TCG AAC CGA CCA GAA TCA TGC AAG TGC GTA AGA TAG TCG CGG GCC GGG 3' and corresponding bottom sequence 5' TCG ACC CGG CCC GCG ACT ATC TTA CGC ACT TGC ATG ATT CTG GTC GGT 5' were annealed by mixing 1.5 ug of each primer in 20 ul of annealing buffer (10 mM Tris-HCl, 100 mM NaCl and 1 mM EDTA) in an eppendorf tube. The tube was then placed in a beaker with boiling water, and was left until the beaker cooled to RT. The annealed product was then phosphorylated using T4

polynucleotide kinase according to manufacturer's instructions. Next, 100 ng of the S1 aptamer fragment were mixed with 100 ng of dephosphorylated pSK-pAUL26 vector digested with XhoI. Ligation was performed as described above except that plasmids obtained were screened for the desired insert by digestion with HindIII and KpnI. The XhoI site was destroyed during this cloning process and the resulting plasmid containing the PAS of *UL26* as well as the S1 aptamer was named pSK-pAUL26-tag. This plasmid was used as a starting vector for the production of the other S1 aptamer constructs.

The sequence containing the *UL26* PAS was removed by restriction digest using either PstI or BamHI and EcoRI. Then plasmids containing the polyA regions were digested with either PstI for pSK-pAUL24, or with BamHI and EcoRI for pSK-pA-UL38, pSK-pAUL38inv, pSK-pAUL40 and pSK-pAUL40inv to excise the polyA fragments. The digested vectors, which contained the sequence encoding the S1 aptamer, were gel purified using the SIGMA GenElute™ gel extraction kit and dephosphorylated using the Antartic phosphatase. The inserts were ligated with their respective vectors to obtain pSK-pAUL24-tag, pSK-pAUL24inv-tag, pSK-pAUL38-tag, pSK-pAUL38inv-tag, pSK-pAUL40-tag and pSK-pAUL40inv-tag. Ligations were carried out as described above and the plasmids were screened with the same restriction sites used for cloning, either PstI, or EcoRI and BamHI. The orientations of the inserts were determined by DNA sequencing.

3.5 *In vitro* transcription

In vitro transcription was performed using the T3 MEGAscript™ Kit (Ambion). All centrifugations were performed using a microcentrifuge at the indicated temperatures. To generate template DNA, 10ug of each plasmid DNA were linearized by digestion with KpnI, except for pSK-pAUL26-tag, which was linearized with BceAI. The digested template DNA was then resolved by electrophoresis on a 1.2% agarose gel, and purified using the SIGMA GenElute Gel Extraction kit. The DNA obtained was further purified through the addition of 100 ug of proteinase K and 50 ul of a 1% SDS solution to 50 ul of the template, followed by incubation for 30 min at 50°C. To precipitate the DNA, 10 ul of 3M NaAcetate and 200ul of 100% EtOH were added, and samples were left at -20°C overnight. Samples were then centrifuged at 4°C for 30 min at 13,000 rpm and the pellet was washed with 500 ul of 100% EtOH, centrifuged at 13,000 rpm for 5 min and briefly air dried. The DNA pellet was resuspended in 40 ul of dH_2O.

The transcription reaction was assembled at RT following the manufacturer's instructions (Ambion). Reactions with purified DNA template (ranging from 500 ng – 5 ug) were prepared in 120 ul volumes and incubated for 6 hr at 37°C. Phenol-chloroform was used for RNA extraction and overnight RNA precipitation at -20°C was performed using 100% isopropanol. The

precipitated RNA was recovered by centrifugation at 4°C for 15 min at 13,000 rpm. The supernatant was discarded, and the RNA pellet was resuspended in 120 ul of dH2O and stored at -20°C.

3.6 Affinity purification

3.6.1 Validation of RNA binding to beads

In order to confirm that the transcribed S1-tagged RNA binds the streptavidin beads under the established affinity purification system conditions, a small scale affinity purification experiment was carried out in the absence of cell lysates. All steps were done at 4°C unless otherwise specified and all centrifugations were carried out using a microcentrifuge.

To begin with, 100 ul of a 50% streptavidine-agarose bead slurry (Thermo Scientific #20353) were equilibrated three times with 500 ul (10 bead volumes) of binding/wash buffer (50 mM Hepes pH 7.4, 0.1 M NaCl, 10 mM MgCl2). Samples were then centrifuged at 1000 rpm for 1 min to pellet the beads, and after each wash, the supernatant was discarded. Then, 25 ug of RNA were added to the beads in a volume of 300 ul on ice, and incubated with rotation for 1 hr. Samples were then centrifuged at 1000 rpm for 1 min. Supernatants were transferred to fresh tubes and kept at -20°C. Beads were then washed with 1 ml of binding/wash buffer for 2 min with rotation, and centrifuged at 1000 rpm for 1 min. Supernatants were transferred to a fresh tube and kept at -20°C. A second wash was performed with 500 ul of binding/wash buffer. Supernatants were transferred to a fresh tube and kept at -20°C. Elutions were performed by adding 300 ul of binding/wash buffer with 5 mM biotin, on ice, and samples were incubated with rotation for 20 min. Samples were then centrifuged at 1000 rpm for 1 min. These first eluates were transferred to a fresh tube and kept at -20°C. A second elution was performed at RT with 300 ul of 8 M Urea. Samples were incubated with rotation for 5 min, and then centrifuged at 1000 rpm for 1 min. Urea eluates were also stored at -20°C.

Eluates and supernatants were resolved on a denaturing 12% acrylamide urea gel in TBE, which was stained with ethidium bromide (89 mM Tris base, 89 mM Boric acid, 0.2 mM EDTA).

3.6.2 Preparation of lysates for affinity purification

HeLa cell lysates were prepared from mock-infected cells or cells infected with KOS virus at an MOI of 5. Infections were carried out in confluent T175 flasks with ~$2x10^7$ cells per flask. To infect cells, the growth medium was removed and 4 ml of inoculum (prepared in DMEM containing 2% NCS and Pen/Strep) were added to each T175 flask. The flasks were incubated

for 1 hr at 37°C, with shaking every 15 min. The inoculum was then removed, and the cells were washed once with 15 ml PBS. Fresh complete medium (DMEM with 8% FBS, penicillin and streptomycin) were added to each flask (30 ml), and cells were incubated for 5 or 10 hrs.

To prepare the lysates, the growth medium from each flask was removed and the cells were washed with 15 ml of warm PBS (37°C). The wash was discarded and 4 ml of lysis solution containing protease inhibitor (Roche M-lysis Kit) were added to each flask. Cells were lysed for 15-20 min with gentle shaking every 5 min. The lysed cells were collected by tapping the flasks and scraping the lysed cells. The lysed cells were centrifuged at 12500 rpm using the JA-20 rotor for 10 min at 4°C to remove cellular debris. The supernatants were transferred to fresh tubes and stored at -20°C. Thawed lysates were transferred to 12-30 ml Slide-A-Lyzer dialysis cassettes (ThermoScientific), and dialyzed overnight in 2 L of binding/wash buffer at 4°C. Buffer was changed once and samples were dialyzed for another 8 hrs. Dialyzed lysates were stored at -20°C.

3.6.3 Large scale affinity purification

All steps were carried out at 4°C unless otherwise specified. For each experimental condition, dialyzed lysates from 5 T175 flasks were thawed on ice and then centrifuged for 30 min at 12,000 rpm (JA-20 rotor) to remove proteins that precipitated during dialysis. To each lysate sample produced from 5 T175 flasks, 400 ug of avidin (SIGMA) were added and incubated for a minimum of 10 min prior to using the lysates. For each experimental condition tested, 100 ul of streptavidine-agarose beads (Thermo Scientific #20353) were used. Beads were equilibrated by washing with 1 ml (10 bead volumes) of binding/wash buffer (50 mM Hepes pH 7.4, 0.1 M NaCl, 10 mM $MgCl_2$) three times. After each wash, the beads were centrifuged at 1000 rpm for 1 min in a microcentrifuge and the supernatant was discarded. Then, 100 ug of RNA were added to the beads in a volume of 800 ul and incubated with rotation for 1 hr. The samples were then centrifuged at 1000 rpm for 1 min in a microcentrifuge and supernatant containing the unbound RNA was discarded. The beads were then washed with 1 ml of binding/wash buffer for 5 min with rotation, and centrifuged at 1000 rpm for 1 min in a microcentrifuge. This step was done to remove excess RNA that did not bind the beads. The cell lysates from the equivalent of 5 T175 flasks (~25 ml) were added to the beads along with 20 ul of RNase inhibitor (Fermentas) and incubated with rotation for 1 hr. Beads were collected by centrifugation at 300 rpm for 1 min (TTH-750 rotor, Heraeus Multifuge 3SR Plus). The supernatants were kept at -20°C. Beads were washed twice with 1 ml of binding/wash buffer for 5 min with rotation and centrifuged at 1000 rpm for 1 min in a microcentrifuge. Elutions were performed by adding 500 ul of binding/wash buffer with 5 mM biotin to the beads, which

were incubated with rotation for 30 min. Samples were then centrifuged at 1000 rpm for 1 min in a microcentrifuge and these first eluates were stored at -20°C. A second elution was performed at RT in 500 ul of 8 M Urea and samples were incubated with rotation for 10 min. The samples were then centrifuged at 1000 rpm for 1 min, and the urea eluates were stored at -20°C.

3.7 Mass spectrometry

The eluates obtained from the affinity purification procedure were resolved on a large scale 10% polyacrylamide SDS gel. The proteins were stained with silver using the ProteoSilver™ Plus Silver Stain Kit (SIGMA) according to the manufacturer's instruction. The bands of interest were excised and then destained. Destaining was carried out by adding equal volumes of the ProteoSilver Destainers A and B (SIGMA) and incubating for 15-30 min. Gel slices were then washed with 200 ul H_2O OmniSolv three times for 15 min each. To identify the proteins in the gel slices, samples were subjected to analysis by MS using the method described in the work of Charbonneau et al., 2007. Briefly, the wash was removed and 200 ul of 100 mM NH_4HCO_3 were added, and the samples were mixed for 10 min. Then 200 ul of 100% CH_3CN were added and mixed for 10 min. The liquid was removed and 50 ul of 100% CH_3CN were added, and the samples left at RT for 15 min. The CH_3CN was then removed, and the samples were dried in a "speed vac" for 10 min. The samples were then subjected to in-gel tryptic digest, and the tryptic peptides were separated on an Agilent Nanopump system using C18 ZORBAX trap and a SB-C18 ZORBAX 300 reversed-phase column (Agilent Technologies, Inc.). The mass spectra were recorded on a hybrid linear ion trap-triple quadrupole mass spectrometer (Q-Trap; AB Applied Biosystems) equipped with a nanoelectrospray ionization source. MS analysis was carried out at the INRS-Institut-Armand-Frappier MS facilities.

CHAPTER 4: RESULTS

4.1 Evaluation of the steady-state levels of polyadenylation factors during infection by HSV-1

4.1.1 Cell lysate preparation

As reported by Cook and Coen (1996), the short *UL24* transcripts of HSV-1 are expressed with early kinetics, while long transcripts are expressed with late kinetics. In order to understand the change in the polyadenylation pattern of the *UL24* transcripts during infection, we monitored the steady-state levels of different subunits of the cellular polyadenylation machinery. Whole cell lysates were prepared from 60 mm dishes of HeLa and Vero cells using a RIPA lysis buffer with high salt (500 mM NaCl). Vero cells were chosen as they are commonly used for studies involving HSV-1 gene expression. In addition, HeLa cells were used for these studies because humans are the natural hosts for HSV-1. Lysates were collected from mock-infected cells as well as from cells infected at an MOI of 5 for 3, 6, 9, 12 and 16 hrs. These time points allowed us to cover the entire temporal cascade of viral genes expressed, and evaluate the levels of polyadenylation factors present throughout the infection.

4.1.2 Western blots for polyadenylation factors

Cell lysates obtained were evaluated through Western blotting using commercially available antibodies against subunits from different components of the 3'end processing machinery. Several commercial antibodies against the different polyadenylation factors were tested; CPSF-1 (CPSF-160), CPSF-2 (CPSF-100), CPSF-3 (CPSF-73), CPSF-4 (CPSF-30), CstF-1 (CstF-50), CstF-2 (CstF-64), CstF-3 (CstF-77), CPSF-5 (CFI_m-25kDa), CPSF-6 (CFI_m-68kDa) and CPSF-7 (CFI_m-59kDa). However for many of these factors, the available antibodies were not specific or showed a large amount of background binding such that it was not possible to assess with certainty which band corresponded to the subunit studied (data not shown). Nevertheless, we were able to evaluate one subunit each from CPSF, CstF and CFI_m (Fig. 11-13). The membranes were also probed for Y-tubulin as a loading control. Results shown are representative of experiments that were performed at least three times.

4.1.2.1 CPSF

CPSF is the largest component of the polyadenylation machinery and interacts with several elements in the 3' UTR. The subunit CPSF-100, which is thought to be a scaffolding protein, was tested (Fig. 11). Though slight fluctuations in band intensities were noted, namely between mock infected lysates and lysates at 3 hpi, no reproducible change was observed in the steady-state levels of CPSF-100 up to 12 hpi in either HeLa or Vero cells. Also, there appeared to be a slight decrease in steady-state levels at 16 hpi in Vero cells; however, this decrease occurred well after the levels of short *UL24* transcripts are reduced (10 hpi). Therefore, the decrease in steady-state levels of CPSF-100 cannot be the cause of the reduced usage of the *UL24* PAS.

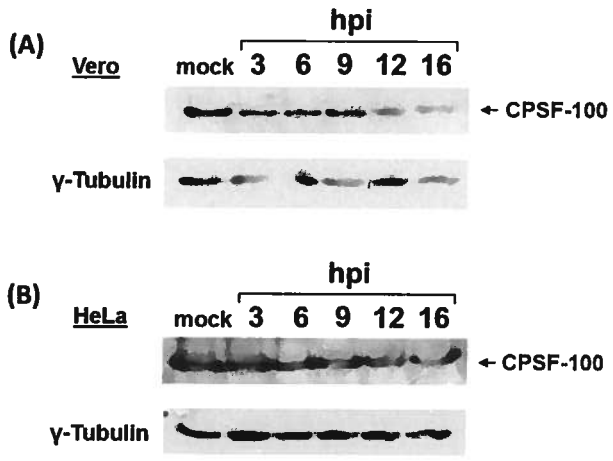

Fig. 11. Levels of CPSF-100 during HSV-1 infection. Western blot analysis of cell lysates from Vero cells (A) and HeLa cells (B) that were either mock-infected or infected for the indicated times. Upper panels show levels of CPSF-100 protein and lower panels show the results for the respective loading control, γ-tubulin.

4.1.2.2 CstF

Next, a component of the CstF factor, which is important for efficient binding and polyadenylation specificity, was evaluated (Fig. 12). Changes in steady-state levels of CstF have previously been shown to be implicated in alternative polyadenylation during B-cell maturation. The subunit tested was CstF-64; it is the subunit responsible for binding to the G/U-rich DSE, and it helps guide the polyadenylation machinery to the correct site on the transcripts. Experiment was repeated several times and though some fluctuations were observed in expression levels, namely HeLa mock-infected vs. 3hpi, the results obtained show no reproducible change to the steady-state levels of this subunit during the infection for either HeLa or Vero cells.

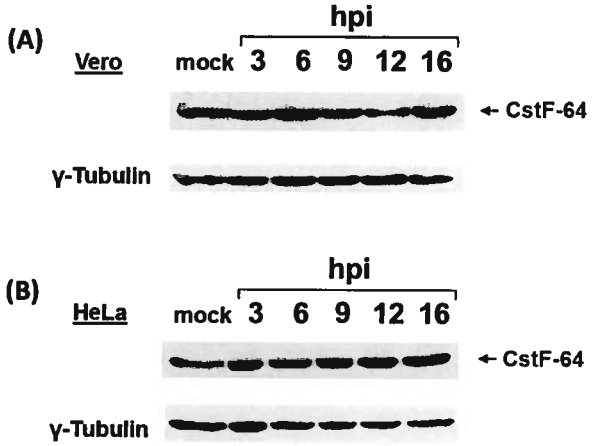

Fig. 12. Levels of CstF-64 during HSV-1 infection. Western blot analysis of cell lysates from Vero cells (A) and HeLa cells (B) either mock-infected or infected for the indicated times. Upper panels show levels of CstF-64 protein and lower panels show the results for the respective loading control, γ-tubulin.

4.1.2.3 CFI_m

The 25kDa-subunit of the CFI_m factor, required for transcript cleavage, was also evaluated (Fig. 13). It interacts with both the PAP and PABP, and binds RNA. Though some fluctuation was seen in band intensities, namely between mock infected HeLa lysates and lysates at 3 hpi, the results obtained show no reproducible change in the steady-state levels of this subunit during infection of either Vero or HeLa cells.

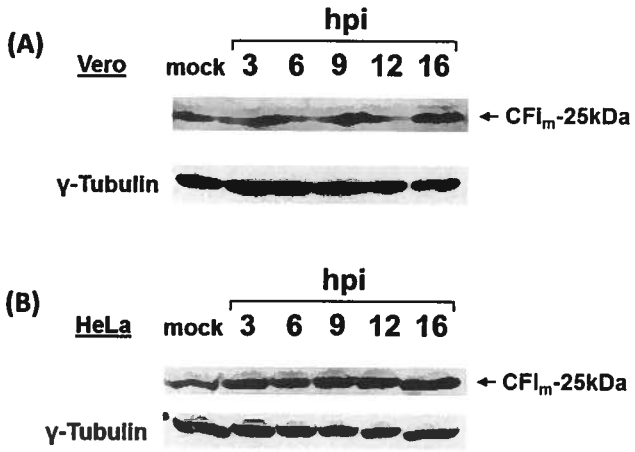

Fig. 13. Levels of CFI_m-25 during HSV-1 infection. Western blot analysis of cell lysates from Vero cells (A) and HeLa cells (B) either mock-infected or infected for the indicated times. Upper panels show levels of CFI_m-25 protein and lower panels show the results for the respective loading control, γ-tubulin.

In summary, the alternative polyadenylation observed leading to expression of the short *UL24* transcripts at early times and expression of the long transcripts at late times during infection cannot be attributed to changes in the steady-state levels of the factors evaluated, namely CPSF-100, CstF-64 or CFI_m-25.

4.2 Affinity purification of RNA binding protein with streptavidin

An affinity purification system was set-up to isolate factors that interact with regions containing the polyadenylation signals of different alternatively polyadenylated viral transcripts.

4.2.1 Plasmid construction

In order to study the polyadenylation of the viral transcripts for *UL24*, *UL26*, *UL38* and *UL40*, the regions containing the PAS of each of the four genes were cloned into the plasmid pBluescript SK+ (Stratagene) (Fig. 14 and 15). Each plasmid also contained the 44 nt sequence for the aptamer that recognizes streptavidin (S1-tag), which was inserted by ligation into the nearby XhoI restriction site. The plasmids served as template DNA for large scale *in vitro* transcription reactions.

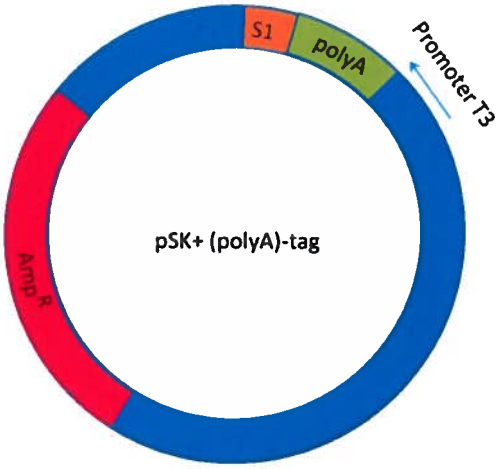

Fig. 14. Illustration of polyA-S1 tag plasmid constructs. pBluescript SK+ was used as the starting vector, it contains the amplicilin resistance gene. Fragments containing the PAS (green) of the different genes were inserted into the multiple cloning site of pSK+, downstream of the T3 promoter, except for the negative controls which were inserted in the inverse orientation. The S1 tag (orange) was cloned into the nearby XhoI restriction site.

***UL24* PAS**

5′ UGC CAA CGC CAG ACG CCG GUC CGC UGU GCC AUC GNU CCC CUU CAU CCC ACC CCC AUC UUG UCC CCA AAU AAA ACA AGG UCU GGU AGU UAG GAC AAC GAC CGC AGU UCU CGU GUG UUA UUG UCG CUC UCC GCC UCU CGC AGA 3′

***UL26* PAS**

5′ CGA UUU GUU CGU CUC UCA GAU GAU GGG GGC CCG CUG AUU CGC CCC GGU CUU UGG UAC CAU GGG AUG UCU UAC UGU AUA UCU UUU UAA AUA AAC CAG GUA AUA CCA AAU AAG ACC CAU UGG UGU AUG UUC UUU UUU UUU UUU UAU UGG GAG GGG CGG GUA GGC GGG UAG CUU UAC AAU GCA AAA GCC UUU GAC GUG GAG GAA GG 3′

***UL38* PAS**

5′ UGA GCG UAG CAA ACG CCC CGC CCA CAC AAC GCU CCG CCC CCA ACC CCU UCC CCG CUG UCA CUC GUU GUU CGU UGA CCC GGG CGU CCG CCA AAU AAA GCC ACU GAA ACC CGA AAC GCG AGU GUU GUA ACG UCC UUU GGG CGG GAG GAA GCC ACA AAA UGC 3′

***UL40* PAS**

5′ AGU GUC GCA GCA CCU CCU ACG CCG GGG CGG UCG UCA ACG AUC UGU GAG ***G***GU CGC GGC GCG CUU CUA CCC GUG UUU GCC CAU AAU AAA CCU CUG AAC CAA ACU UUG GGU CUC AUU GUG AUU CUU GUC AGG GAC G 3′

U-rich upstream element

PAS

GU-rich downstream element

G-rich downstream element

Fig. 15. Sequences containing the PAS of all four genes cloned into pBluescript SK+. Sequences shown highlight location of the *cis*-acting elements.

4.2.2 *In vitro* RNA transcription

Transcripts were generated using the T3 MEGAscript™ Kit (Ambion). The RNA synthesized consisted of the various 3' UTR with the S1-tag fused at the 3'end. The S1-tag was placed at the 3'end of the sequences containing the PAS inserted to ensure that only the full length transcripts bind to the beads, though we cannot rule out the possibility that partially degraded RNA fragments containing the tag were also retained. Template DNA was prepared for the *in vitro* transcription reactions by linearization through digestion with the appropriate restriction enzyme. The amount of template DNA (i.e. 0.5 ug – 5ug), the temperature of the transcription reaction (i.e. 20, 25, 30 or 37 °C), as well as the incubation times (i.e. 4, 5 or 6 hrs) were optimized (data not shown). Each transcription reaction yielded 100-250 ug of RNA. An aliquot of the RNA produced for each reaction was subjected to denaturing polyacrylamide gel electrophoresis, and a band corresponding to the correct molecular mass was observed. We did however see higher molecular mass bands that likely represented RNA that was extensively folded, as well as smaller bands possibly resulting from nicks present in the template DNA used.

4.2.3 Validation of the S1 tag affinity purification system in the presence of the RNA ligand only

We wanted to confirm that in the RNAs we synthesized *in vitro*, the S1-tag was transcribed and was available to bind streptavidin. Using the same binding conditions chosen for the affinity purification, we incubated the beads in the presence of the RNA generated, in the absence of lysate. After several washes, a 20 min elution step using 5 mM biotin was carried out followed by a 5 min elution with 8 M urea. The eluates were resolved by denaturing polyacrylamide gel electrophoresis, and the RNA was stained with ethidium bromide (Fig. 16). We were able to obtain specific elution of the RNAs using biotin (fig. 16-EB lane). Also, a more prominent band was seen in the 8 M urea elution (fig. 16-EU); indicating that a significant portion of ligand remained bound after the short 20 min biotin elution, and could be displaced by addition of urea.

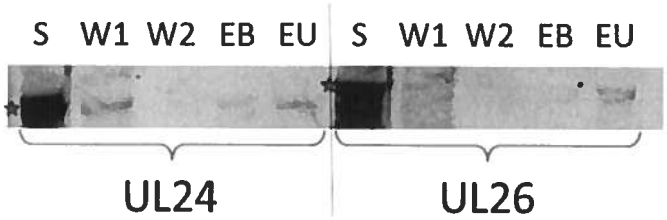

Fig. 16. RNA gel showing affinity purification with ligand only. Left panel showing results for pAUL24-tag RNA and right panel showing pAUL26-tag RNA results. S, W1, W2, EB and EU, show the supernatant after RNA addition, the first wash, the second wash, the biotin elution and the urea elution, respectively. Stars indicate RNA bands with the expected sizes of transcripts.

4.2.4 Lysate preparation

At first, a mild lysis approach using the dounce homogenizer (B-type pestle) described by Sopta et. al for the preparation of lysates for isolating RNA-binding proteins was used (Sopta, Carthew, and Greenblatt, 1985), however it was not efficient in lysing our HeLa cells. Instead, lysates were prepared using the Roche complete M-lysis kit for mild lysis conditions. Lysates were collected from uninfected cells as well as cells infected at an MOI of 5, for 5 or 10 hrs. These times were chosen because they represent the times corresponding to early and late viral transcript expression. Lysates from approximately $1x10^8$ cells were incubated with each affinity matrix corresponding to a different RNA. Because the lysis buffer used in this kit was not compatible with affinity purification system, a dialysis step was required to exchange the lysis buffer for the binding buffer. Some proteins present in the cell lysates were lost through precipitation as detected visually at the end of the dialysis steps. This precipitation of proteins was not observed in a previous smaller scale dialysis under the same conditions. Also, attempts were made to add certain components to the binding buffer that would be expected to help prevent RNA cleavage, such as TPEN (8 uM), a zinc chelator that inhibits CPSF-73 endonuclease activity (Ryan, Calvo, and Manley, 2004), or enhance protein stability (0.1 mM DTT and 10% glycerol); however, the affinity purification system did not tolerate this modified binding buffer.

4.2.5 Affinity purification of RNA-binding proteins

In this study, only the PAS of *UL24* and *UL26* were tested to first validate the affinity purification system. The steps involved in the affinity purification system are briefly described here and are outlined in Fig. 17. The streptavidin beads were equilibrated in binding buffer and incubated at 4°C for 1 hr with the transcripts generated *in vitro* containing the 3' UTRs of either *UL24* or *UL26* or as negative control for binding of polyadenylation specific factors, the inverse *UL24*

3'UTR sequence, each fused to the nucleotide sequence corresponding to the S1-tag aptamer. The beads were washed, lysates were added and the mixture was incubated at 4°C for 1 hr. Beads were washed again and bound proteins were eluted using biotin. A second elution was performed using 8M urea.

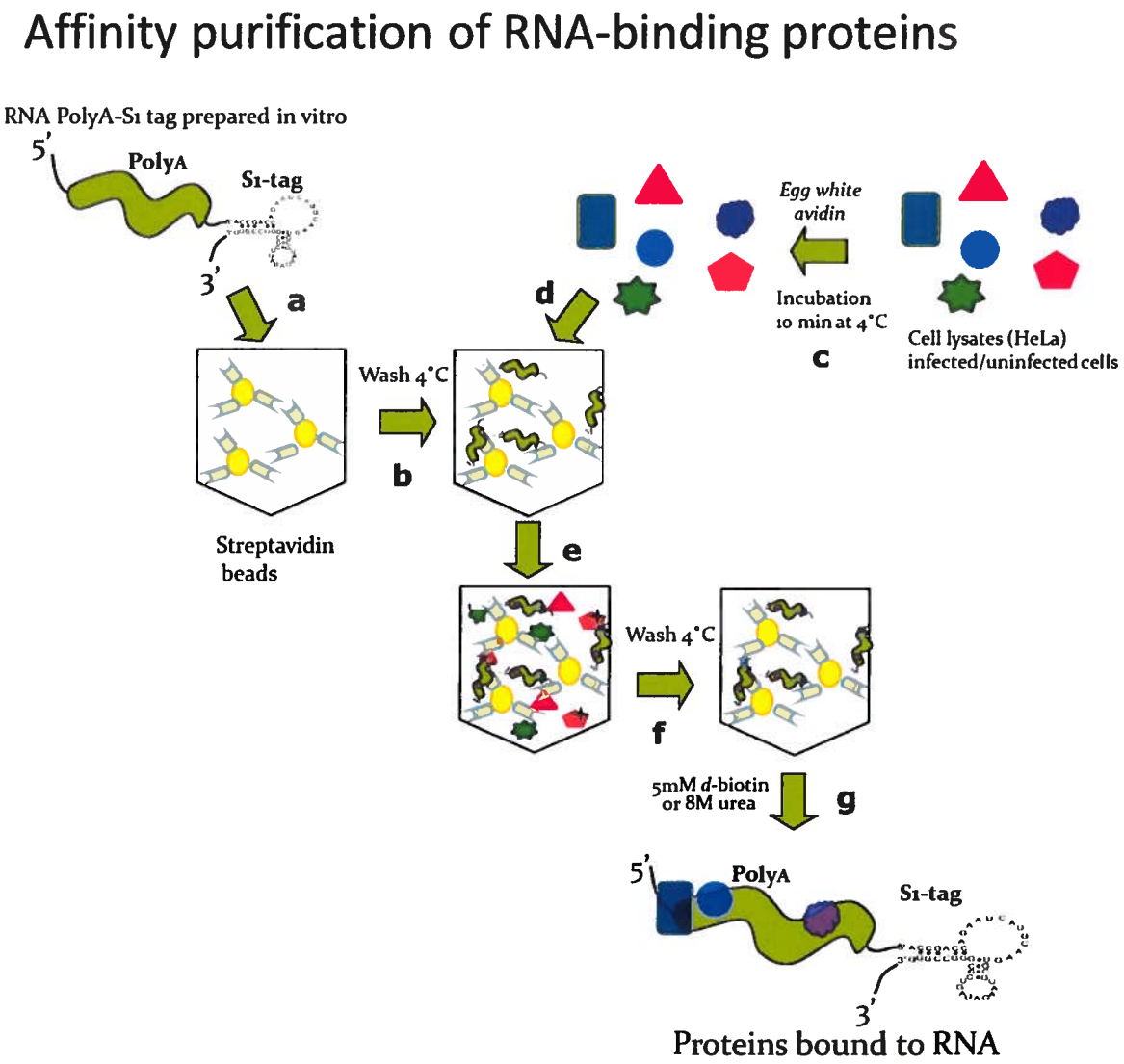

Fig. 17. Flow chart of affinity purification system. Outline of affinity purification steps: (a) addition of transcribed RNA to equilibrated streptavidin beads and incubation, (b) wash to remove excess, unbound RNA, (c) HeLa lysate pre-incubated with egg white avidin, (d) addition of HeLa lysates to RNA-bound beads, (e) incubation, (f) washes to remove unbound protein and (g) elutions to obtain protein-bound-RNA product.

Proteins present in eluates were resolved on a 10% polyacrylamide denaturing gel with SDS and stained with silver (Fig. 18 and 19). Specific bands were detected in the biotin elution. For both the biotin and the urea elution, bands were more prominent in the samples collected at 10 hpi; however, this may simply reflect the higher amount of proteins in the 10 hpi samples. This is likely due, at least in part to an error leading to a third of the mock lysate being added to the 10 hpi lysate. In addition, there seemed to be more protein in the eluates from the pAUL26-tag RNA samples. Bands to be excised were chosen based on the relative intensity of the band in the different samples at the same time point; bands that were more intense in the samples containing the pAUL24 and/or pAUL26 as compared to the pAUL24inv were selected for analysis. Selected bands were excised from the gel and sent for analysis by liquid chromatography - mass spectrometry (LC-MS) (Fig. 18 and 19).

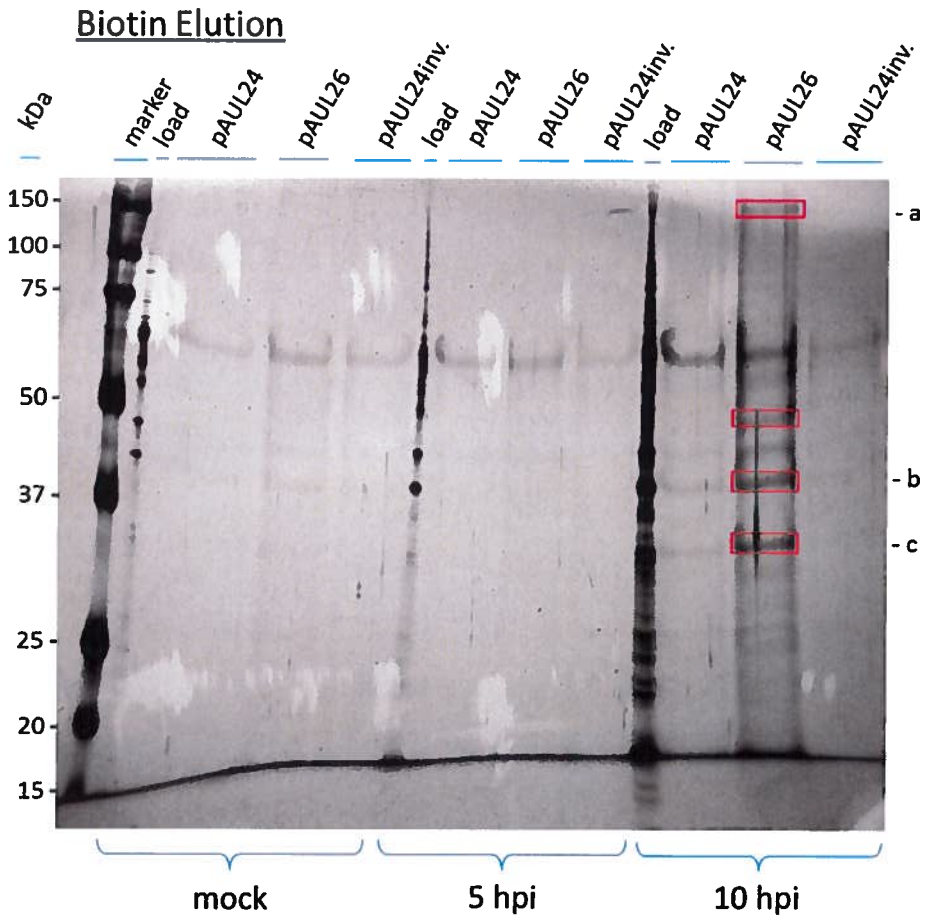

Fig. 18. Silver stained gels showing biotin eluates. Shown are the eluates from mock-infected cells and cells infected for 5 and 10 hrs. For each time point, the experiment was performed with the RNA from the pAUL24-tag, pAUL26-tag and the negative control pAUL24inv-tag. Bands excised are highlighted by the red boxes. Bands associated with letters gave hits in the LC-MS analysis. Positions of molecular mass marker are indicated to the left of the panel. Shown in the "load" lanes, are small portions of lysate samples used.

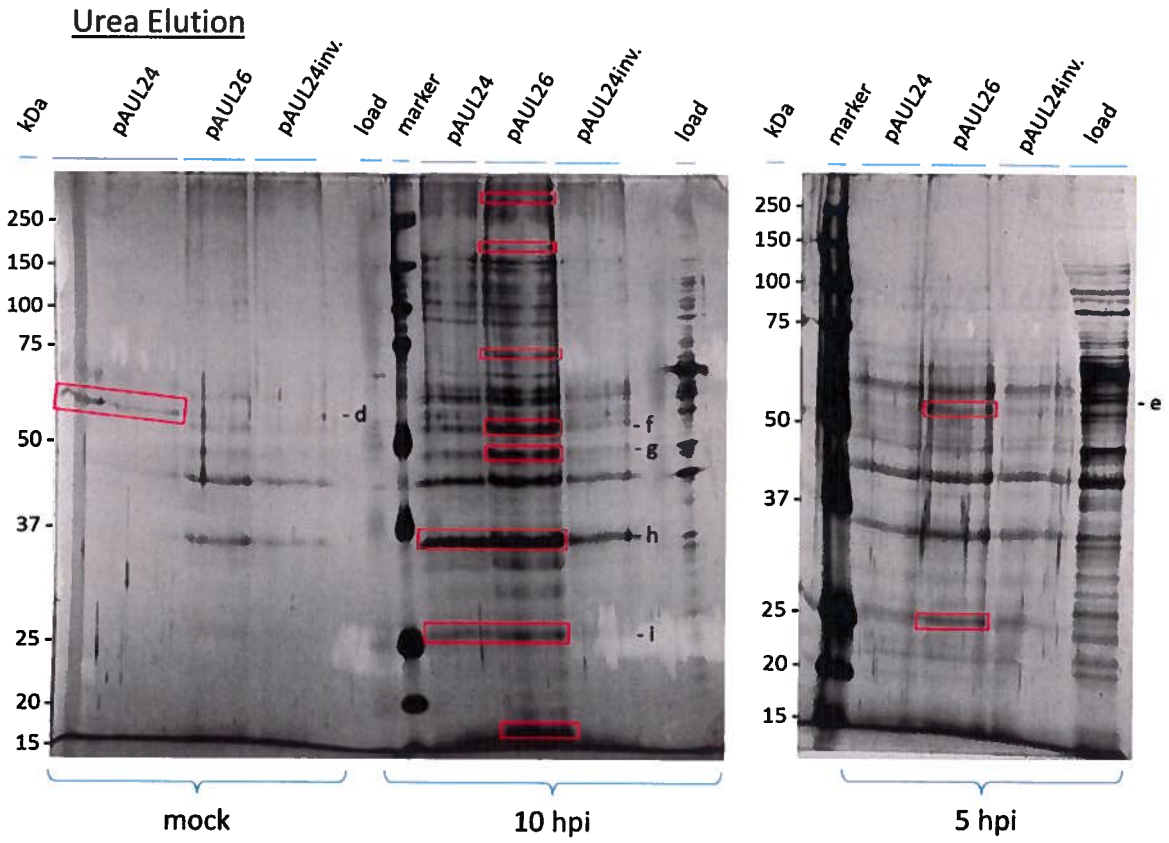

Fig. 19. Silver stained gels showing urea eluates. Shown are the eluates from mock-infected cells and cells infected for 5 and 10 hrs. For each time point, the experiment was performed with the RNA from the pAUL24-tag, pAUL26-tag and the negative control pAUL24inv-tag. Bands excised are highlighted by the red boxes. Bands associated with letters gave hits in the LC-MS analysis. Positions of molecular mass marker are indicated to the left of the panel. Shown in the "load" lanes, are small portions of lysate samples used.

4.2.6 Analysis by LC-MS

Bands were selected, excised and sent for LC-MS analysis (Fig. 18 and 19). Proteins were digested with trypsin, tryptic peptides were separated using reverse-phase liquid chromatography and mass spectra were recorded on a hybrid linear ion trap-triple quadrupole mass spectrometer. Of the 15 bands excised, eight bands gave hits with protein scores of 43 or more. Scores of 43 or more indicate that the probability that the observed match is a random event is $p \leq 0.05$. The proteins identified from the hits obtained were: actin, chaperonin, dynein, enolase, HSP-27, leucine-rich repeat serine/threonine-protein kinase 2 (LRRK2), myosin, nuclear factor 45 (NF45), transglutaminase, translation elongation factor 1 (eEF1A), tubulin, and some unknown proteins (Table 1). Though the hit score was only 40, elongation factor RNA polymerase II (ELL) was also included in the table due to its role in transcription. In addition, there were hits for various isoforms of the contaminant keratin, which were not included in the table.

Most of the proteins identified migrated at levels close to their expected molecular masses. A couple of proteins were identified from bands migrating at lower molecular weights than the predicted molecular weight of the protein; LRRK2, prostate-specific transglutaminase 4 and dynein. This discrepancy could be caused by degradation of the proteins at some point in the purification process. Unexpectedly, the elongation factor RNA pol II was identified from a band corresponding to ~125kDa while its predicted weight is 80kDa.

Certain proteins were dismissed as being contaminants due to their abundance in the cell or due to their known properties and the likelihood that they bind to the matrix through non-specific interactions. These include actin, tubulin, myosin, dynein, enolase, transglutaminase and HSP-27. Surprisingly, we did not identify any polyadenylation factors, although the presence of CSPF-100 and CstF-64 in the starting lysates was confirmed by Western blot (data not shown).

Of particular interest were proteins with known nucleic acid binding properties. These included ELL, NF45 and eEF1A. Using the programs PROSCAN and InterProScan, we found that LRRK2 has a leucine zipper motif, as well as a signature of the regulator of chromosome condensation RCC1, and therefore also has potential nucleic acid binding properties.

Motif searches were also performed on the unknown protein amino acid sequences but no nucleic acid binding motifs were identified in those hits. Other motifs with various associated functions were also identified during the process such as ankyrin, WD40, aramadillo-like helical and Leucine-rich motifs. Most of these are involved in protein-protein interactions, but provided no specific evidence in support of any gene regulatory function.

Sample	Elution	Protein hits	Number of peptides identified	Protein score	Molecular mass of excised band (kDa)	Molecular mass of predicted protein (kDa)
a	Biotin/UL26	elongation factor RNA polymerase II	2	40	~125	80
b	Biotin/UL26	mutant beta-actin (beta'-actin) alpha-actin gamma-actin actin, beta-like 2 *similar to* actin, gamma 1 actin-like protein actin-like protein	4 4 3 3 1 6 2	208 187 156 150 86 63 46	~40	42
c	Biotin/UL26	Unnamed	3	46	~32	-
d	Urea/ UL24	hCG2003116, isoform CRA_a	3	49	~57	-
e	Urea/ UL26	Unnamed	3	51	~ 60	-
		Leucine-rich repeat serine/threonine-protein kinase 2	4	50		286
		KIAA1170	3	45		-
f	Urea/ UL26	tubulin, beta 5 tubulin, beta, 2 Unnamed protein (tubulin motifs) alpha-tubulin tubulin, alpha 1B hCG27371-tubulin hCG1747788-tubulin	9 7 7 7 7 2 1	637 507 501 345 343 87 57	~52	48-57
		chaperonin containing TCP1 subunit 2	1	53		60
g	Urea/ UL26	2-phosphopyruvate-hydratase alpha-enolase; carbonate dehydratase Unnamed protein product (enolase) neurone-specific enolase	3 1 1	202 97 45	~47	48
		glucocorticoid receptor AF-1 specific elongation factor *similar to* eukaryotic translation elongation factor 1 alpha 1 isoform 4	3 3	146 118		50
h	Urea/ UL24 & UL26	Actin, gamma 1 propeptide Gamma-actin Actin, beta-like 2 Actin-like protein *similar to* actin, gamma 1	9 7 5 2 1	363 264 151 92 50	~35	42
		hCG2018327	3	52		-
		KIAA0799 (myosin motif identified)	5	52		variable
		NF45 protein	1	47		45
		dynein, axonemal, heavy chain 3	3	43		530
i	Urea/ UL24 & UL26	heat shock protein 27	2	87	~26	27
		prostate-specific transglutaminase 4	3	50		77

Table 1. Summary of hits from affinity purification LC-MS analysis. List of proteins that were identified in the biotin or the urea eluates of the RNA affinity matrices.

CHAPTER 5: DISCUSSION

5.1 Impact of HSV-1 infection on steady-state levels of polyadenylation factors

In order to determine if the switch from short to long *UL24* transcripts is due to changes in the steady-levels of polyadenylation factors over the course of HSV-1 infection, the levels of various cellular polyadenylation factors present in the cell lysates at different times post-infection were assessed by Western blotting.

Western blots performed on the different polyadenylation factors tested showed no reproducible changes in steady-state levels during the first 12 hours of infection. This suggests that the shift in PAS usage is not caused by a change in the steady-state levels of these factors; however, there are still several polyadenylation factors that could not be evaluated due to the lack of specificity of the commercially available antibodies. Thus, we cannot rule out the possibility that changes in the steady-state levels of some of the untested factors are involved in alternative polyadenylation during infection. Also, in order to more accurately assess the variations in steady-state levels of the polyadenylation factors, and to properly quantify any possible differences, we would have to load different volumes of lysate and use them as standards to establish the relative differences during the infection. Moreover, there are however several published antibodies against various polyadenylation factors that could be used instead of the commercial antibodies tested in this work (Barabino et al., 1997; de Vries et al., 2000; Jenny, Hauri, and Keller, 1994; Ruegsegger, Blank, and Keller, 1998; Ryan, Calvo, and Manley, 2004).

Even if the steady-state levels of all the factors are unchanged during infection, it could be that the level of association with other proteins or nucleic acids is affected by post-translational modifications. In fact, every factor in the 3'end processing complex is affected by these modifications, see review (Millevoi and Vagner, 2010). For example, certain factors must be phosphorylated to maintain their activity. This is the case for the CFI_m and $CFII_m$ factors, where dephosphorylation abolishes cleavage activity (Ryan, 2007). Another example of regulation through post-translational modification is observed upon accumulation of the phosphorylated form of Pta1p; the symplekin homolog in yeast, which leads to shortened poly(A) tails due to a phosphatase depletion (He and Moore, 2005). Also, hyperphosphorylation of PAP leads to downregulation of its activity and impacts cell growth (Colgan et al., 1996; Zhao and Manley, 1998). Acetylation, another post-translational modification, of the PAP also affects its function since the acetylated form of PAP is unable to associate with its partner CFI_m25 (Shimazu, Horinouchi, and Yoshida, 2007).

Another possible mechanism leading to changes in the PAS processivity is the relocalization of polyadenylation factors rendering them unaccessible to the site of polyadenylation. This is the case for the human cytomegalovirus (HCMV), a beta human herpes virus (Gaddy et al., 2010). During the HCMV infection, the CstF-64 is relocalised to viral replication compartments of the cell, which are sites of transcription and posttranscriptional processing. The presence of CstF-64 at these sites could lead to favouring polyadenylation of the viral transcripts. A change in subcellular localisation is also seen for the PABP during the course of the HSV-1 infection (Dobrikova et al., 2010). It was observed that the total levels of PABP remain unchanged during the infection; however, the cytoplasmic PABP, PABPC1, was seen to accumulate in the nucleus. The authors propose that this is a method employed by the virus to achieve co- and posttranscriptional regulation of mRNA processing and/or nuclear export.

5.2 Affinity purification of RNA binding proteins using a streptavidin-binding RNA tag

In order to assess the level of association of the polyadenylation factors at the different polyadenylation sites during the course of infection, an RNA affinity purification system was developed to identify proteins that interact with the 3'UTRs of viral transcripts. The system was tested on the PAS of *UL24* and *UL26*, which corresponds to the 3'ends of the long and short *UL24* transcripts. We expected to detect more binding of proteins involved in the stimulation of polyadenylation to the 3'UTR of *UL26* than that of *UL24*, since the *UL24* PAS is weaker. We also expected to find proteins that inhibit polyadenylation in the *UL26* eluates at early times during the infection, or in the different *UL24* eluates, since in both cases the PAS are not very efficient.

The affinity purification strategy employed allowed the isolation of several proteins that associate with the tagged RNA. The variability in the apparent protein concentration of the different lysates could be due to several factors. We observed the formation of precipitates during the dialysis step, but the amount appeared to vary between samples. Most of the bands sent for analysis were faint and identification by MS was difficult, as we seemed to be at the limit of detection of the apparatus. Nevertheless, we found that the affinity purification resulted in the identification of several nucleic acid binding proteins, among other types of proteins.

The affinity purification system we used can be optimized at many levels. During the plasmid construction, we took care to include the identifiable cis-acting polyadenylation elements in the 3'UTR we selected for cloning; however, it may be that the region cloned did not include all the elements that allow for efficient association of the various factors and proper formation of the polyadenylation complex.

The binding buffer plays a critical role in these experiments as we are trying to isolate the RNA-protein complex formed on the PAS, which normally undergoes cleavage. Unfortunately, the buffer we originally used, which was designed to minimize cleavage of the RNA as well as protein denaturation, interfered with binding of the S1-tag to the streptavidin beads and could not be used for the affinity purification. Further optimizations would need to be done to assess the conditions for improved protein binding of the polyadenylation factors that would also be compatible with binding of the S1-tag to streptavidin. In our experiments, the lysis buffer used, that of the Roche M-lysis kit, interfered with binding of the S1-tag to the streptavidin beads. As such, dialysis of the lysate into the buffer recommended for binding of the S1-tag was required (Srisawat and Engelke, 2001). This step led to significant loss of proteins due to precipitation; certain polyadenylation factors and binding partners were likely lost in this process. An alternative may be to lyse the cells through sonication directly into the recommended binding buffer. With this approach, the volume of the lysate can also be better controlled allowing for a more concentrated lysate as compared to the lysate obtained with the Roche M-Lysis kit. This was a limitation of the Roche M-Lysis kit, which requires 5 ml lysis buffer for each T175 flask lysed. Another concern is that the content of the commercial buffer is not known, and the lysis process may cause post-translational modifications to the polyadenylation factors which may affect their binding and function as discussed earlier.

5.3 Proteins identified by LC-MS using the affinity purification system

5.3.1 eEF1A

eEF1A was identified with a relatively high MS hit score of 118 and was excised from the pAUL26-tag RNA 10 hpi lane. eEF1A is a multifunctional protein. It plays a role in protein translation, where it assists in the nuclear export of mature tRNA to the cytoplasm and delivery to the ribosome (Grosshans, Hurt, and Simos, 2000; Negrutskii and El'skaya, 1998; Reed, Wastney, and Yang, 1994). It also plays a role in the nuclear export of the PABP1 in an RNA pol II transcription dependant fashion (Khacho et al., 2008). Moreover, it plays an important role in translation fidelity as mutations in its GTP-binding motif shows decreased translation fidelity (Carr-Schmid et al., 1999). Because levels of translation of viral proteins are high during infection, it is possible that the association of eEF1A with viral transcripts is important to ensure the fidelity of translation.

5.3.2 NF45

NF45 is another nucleic acid binding protein that was identified (MS hit score of 47). It was isolated from the pAUL24-tag RNA and pAUL26-tag RNA sample lanes corresponding to 10 hpi, and bands were pooled and sent for LC-MS analysis. NF45, also known as the interleukin enhancer-binding factor 2, is a highly conserved transcriptional activator that forms a complex with the comparatively well-studied factors NF90 and NF110 (Parrott et al., 2005; Zhao et al., 2005). All three components contain a dimerization zinc finger nucleic acid-binding domain able to bind dsRNA (Meagher et al., 1999). There are many reports of NF90 interacting with double-stranded and structured single-stranded RNA (Liao, Kobayashi, and Mathews, 1998; Parrott et al., 2005; Reichman and Mathews, 2003). NF90 was also shown to bind 3'UTRs and aids in mRNA stability (Shi et al., 2005; Shim et al., 2002). This result further confirms that the strategy used allowed us to isolate RNA-binding proteins. The properties of NF45 suggest a role in post-transcriptional regulation especially that its partner, NF90, binds 3' UTRs and plays a role in transcriptional regulation.

5.3.3 ELL

ELL was also identified through the MS analysis. The band was excised from the biotin elution in the pAUL26-tag RNA sample lane. It should be noted that its hit score was 40, which was slightly below the cut-off score of 43. Also, the protein was identified from a gel slice corresponding to a molecular mass of ~125kDa, which is larger than the expected size of ELL, 80kDa; however, ELL is very interesting as it is one of the factors that regulate the activity of RNA pol II during the transcription elongation phase (Shilatifard et al., 1997a). Although the specificity of this hit is questionable, we will discuss it here as it is an example of the type of factor we hypothesized we would identify in this study.

ELL assists in elongation by supressing pausing of RNA pol II transcription complex (Shilatifard et al., 1997b). In the context of infection, supressing the transcriptional pausing could lead to transcription past the first PAS, that of *UL24*, thus favouring the generation of the long *UL24* transcripts. In fact, in B-cells, along with increased loading of the CstF-64 subunit during usage of the weak PAS of the IgM transcript, ELL2 and RNA pol II were also found to bind more tightly to the transcription start site (Martincic et al., 2009). ELL2 is a member of the RNA pol II elongation factor with 66% similarity to ELL (Shilatifard et al., 1997a). Moreover, CstF-64 and ELL2 binding is dependent on the phosphorylation of the CTD of RNA pol II and it is proposed that ELL2 promotes CstF-64 binding to the RNA pol II (Cammas et al., 2008). Unfortunately, it seems that we were at the limit of the detection for the analysis of the samples, however, with

some improvements to this technique, it is should be possible to identify with confidence more proteins like this one which may have been slightly below the cut-off score.

These results show that it is possible to identify proteins that bind viral PAS using this streptavidin affinity purification system, although further optimizations are required. Thus, this technique will be useful in the identification of factors involved in alternative polyadenylation.

CHAPTER 6: PERSPECTIVES

No reproducible change in the steady-state levels of the three subunits of polyadenylation factors – CPSF-100, CstF-64 and CFI_m-25 – was observed during the switch in the expression of short to long transcripts during the infection. It is possible that alternative polyadenylation is caused by changes in the steady-state levels of some of the other subunits. And though the steady-state levels may remain unchanged, the localisation, the levels of association with the RNA or the activity of certain factors may be affected during the course of HSV-1 infection and would have to be evaluated.

Using LC-MS, the RNA affinity purification system developed allowed us to identify eEF1A and NF45; both are nucleic acid binding proteins. eEF1A plays a role in translation fidelity and also assists in the export of mature tRNA, and NF45 is a transcriptional activator. Nonetheless, we appeared to be at the limit of detection of the MS being that about half the excised bands could not be identified, and so a larger scale preparation will help identify more proteins interacting with the affinity matrix. Also, the results presented in this study must be further validated. One approach that could be useful in analysing RNA-protein interactions is the gel mobility shift assay where the migration of RNA is monitored through radioactive labelling to assess if the RNA is protein bound or free. RNA bound to proteins, but not free RNA, would show changes in migration during gel electrophoresis. This would validate the RNA protein interactions identified by MS.

This work will help in understanding the mechanism involved in genetic regulation by alternative polyadenylation. Having validated the system using the *UL24* and *UL26* PAS, future work will include similar studies on other alternatively polyadenylated genes of HSV-1, namely *UL38*. Furthermore, using HSV-1 as a model, this approach brings us to a better understanding of genetic regulation by alternative polyadenylation in eukaryotes. Moreover, a better understanding of the genetic regulation of HSV-1 will help in the development of safe and efficient HSV-1 vectors for use in gene therapy.

ANNEXE 1: RÉSUMÉ FRANÇAIS

IDENTIFICATION DES PROTÉINES QUI INTERVIENNENT DANS LA POLYADÉNLYATION ALTERNATIVE LORS DE L'INFECTION PAR LE VIRUS HERPÈS SIMPLEX 1.

INTRODUCTION

Les Herpèsvirus

L'ordre des *Herpesvirales* comprend trois familles: *Herpesviridae*, *Alloherpesviridae* et *Malacoherpesviridae*. Les virus herpès humains font partie des *Herpesviridae*. Cette famille est divisée en trois sous-familles: *alphaherpesvirinae*, *betaherpesvirinae* et *gamaherpesvirinae* (Davison et al., 2009). Le virus herpès simplex I (VHS-1) fait partie de la sous-famille des *alphaherpesvirinae* et est utilisé comme prototype pour l'étude de cette famille.

Pathogenèse du VHS-1

Le VHS-1 infecte 18 à 35% des enfants, et près de 90% des adultes dans le monde sont atteint du VHS-1 (revue de Desselberger, 1998). De plus, près de 1,500 nouveau-nées sont infectés par le VHS-1 chaque année aux États-Unis (revue de Kimberlin, 2004). Comme tous membres des *alphaherpesvirinae*, le VHS-1 est neurotropique. La réplication du virus au niveau des muqueuses cause des lésions au niveau de la peau, elle peut aussi causer des kératoconjonctivites ainsi que des encéphalites. De plus, chez les individus immunodéprimés le VHS-1 peut se disséminer et infecter la peau, la voie respiratoire, l'œsophage et la voie gastro-intestinale (revues de Jacobs, Breakefield, and Fraefel, 1999; Knipe and Cliffe, 2008).

L'infection par le VHS-1 comprend une phase lytique (réplication productive) et une phase latente (réplication non-productive). Après l'entrée du virus au niveau des muqueuses, le virus infecte les cellules épithéliales où se produisent de nouveaux virions, qui eux pourront infecter d'autres cellules épithéliales. Le virus produit pourra infecter les neurones qui innervent cette région et la capside sera alors transportée au corps neuronal. L'ADN viral persistera dans le noyau sous forme d'épisome et le virus entre alors en latence (revue de Knipe and Cliffe, 2008). La réactivation du virus et l'expression des gènes lytiques sont provoqués par des stimulis tels que l'exposition aux rayons UVs ou le stress qui entraineront la phase productive et l'apparition à nouveau des symptômes de l'infection (Desselberger, 1998).

Le génome du VHS-1 et la régulation génétique

Le VHS-1 présente un génome large, de 152 kpb, linéaire et à ADN double brin (revue de Pellett and Roizman, 2007). Il contient une région unique longue et une région unique courte, et ces deux régions présentent des séquences répétées à leurs extrémités (Wadsworth, Jacob, and Roizman, 1975). Le génome code pour 84 cadres de lecture ouverts (ORF) orientés dans les

deux sens, et qui se chevauchent; indiquant une forte densité d'information génétique (revue de Rajcani, Andrea, and Ingeborg, 2004; Roizman, 1996). Les gènes sont classés selon leur cinétique de transcription, une cascade temporelle bien régulée de gènes immédiats-précoces (IE), précoces (E) et tardifs (L) (Honess and Roizman, 1974).

Pour mieux comprendre cette cascade d'expression, plusieurs travaux de recherche ont étudiés les régions promotrices des trois classes de gènes. Tous les gènes ont un élément TATA, par contre, les gènes IE ont aussi l'élément TAATGARAT. La protéine du tégument VP16 forme un complexe avec les facteurs Oct-1 et HCF qui se lie à la séquence TAATGARAT pour activer l'expression des gènes IE (Hughes, La Boissiere, and O'Hare, 1999). De cette manière, les gènes IE ICP4, ICP0, ICP27 et ICP47 peuvent être exprimés, à l'aide de l'ARN polymérase II (ARN pol II) sans la synthèse de novo de protéines virales (revue de Weir, 2001). Ces protéines ICP4, ICP0, ICP27 et ICP22 stimuleront par la suite l'expression des autres classes de gènes viraux (revue de Wagner and Bloom, 1997).

La protéine ICP27 est essentielle pour la réplication lytique et elle est impliquée dans la transition entre l'expression des gènes E et L. Elle joue un rôle dans la régulation transcriptionnelle et post-transcriptionelle des gènes cellulaires et viraux. ICP27 intervient dans la sélection du site de polyadénylation (PAS) et le traitement en 3' d'ARNm (McGregor et al., 1996), soit dans l' inhibition de l'épissage de l'ARNm (Bryant et al., 2001), ou dans la liaison de l'ARN (Ingram et al., 1996) et dans la stimulation du transport ARN sans introns du noyau au cytoplasme (Johnson and Sandri-Goldin, 2009; Phelan, Dunlop, and Clements, 1996).

Les gènes E codent pour les protéines nécessaires pour la réplication de l'ADN viral (Crute et al., 1989; Gottlieb et al., 1990; Olivo, Nelson, and Challberg, 1988; Weller et al., 1983). Le VHS-1 code pour son propre polymérase d'ADN. Le modèle de réplication généralement accepté est celui du cercle roulant ou « rolling circle » (Jacob, Morse, and Roizman, 1979; Skaliter et al., 1996).

Après le début de la réplication du génome viral, l'expression des gènes E est réduite ou arrêtée, tandis que les gènes L sont exprimés préférentiellement. Les gènes tardifs codent entre autres pour les protéines structurales nécessaires pour l'encapsidation de l'ADN viral et pour l'enveloppement du virion. Un modèle pour expliquer la transition entre l'expression des gènes E et les gènes L, est que les complexes de transcription formés avant et après la réplication virale sont différents, et donc les complexes tardifs se lieraient préférablement aux régions promotrices des gènes L (revue de Wagner, Guzowski, and Singh, 1995). De plus, comme la transcription dans les temps tardifs se fait dans les compartiments de réplication viraux (de Bruyn Kops and Knipe, 1988; Knipe et al., 1987), il est possible que les composantes présentes

dans ces compartiments favorisent l'expression des gènes L (revue de Wagner and Bloom, 1997).

On voit donc que le génome large et dense du VHS-1 requiert un système élaboré d'expression génétique qui nécessite une régulation efficace.

Régulation génétique au niveau post-transcriptionnel chez les eucaryotes

La régulation génétique peut se faire à différents stades dans la synthèse de protéines : l'initiation de la transcription et l'élongation, l'épissage, le traitement 3' et 5', la stabilité de l'ARN, l'export nucléaire-cytoplasmique, et la traduction (Ford, Bagga, and Wilusz, 1997; Vinciguerra and Stutz, 2004; Wickens, Anderson, and Jackson, 1997). Tous ces processus peuvent être régulés pour contrôler l'expression des transcrits et des protéines. Au niveau post-transcriptionnel, ceci peut se faire par l'épissage alternatif (revue de Le Hir and Seraphin, 2008), la polyadénylation alternative, la déadénylation (Mauxion et al., 2009) et l'ajout de la coiffe (Kuersten and Goodwin, 2003).

La polyadénylation alternative est un processus qui fait en sorte qu'un gène peut générer des transcrits pouvant être clivés et polyadénylés de façons différentes, générant des transcrits d'ARNm de longueurs différentes (Fig. A1) (revue de Lutz, 2008). Si un des PAS se trouve dans le cadre de lecture ouvert, la longueur de la protéine sera affectée. Si les PAS alternatifs sont dans la région 3'UTR, la stabilité, la localisation, le transport ainsi que la traduction des transcrits peuvent être affectés (Chao et al., 1999). Des études ont prédit que la polyadénylation alternative est un processus présent chez plus de la moitié des gènes humains (Beaudoing et al., 2000; Tian et al., 2005). Ceci suggère que la diversité des transcrits générée par ce processus est importante et donc l'étude de ces transcrits polyadénylés permettra une meilleure compréhension de la régulation génique dans les eucaryotes.

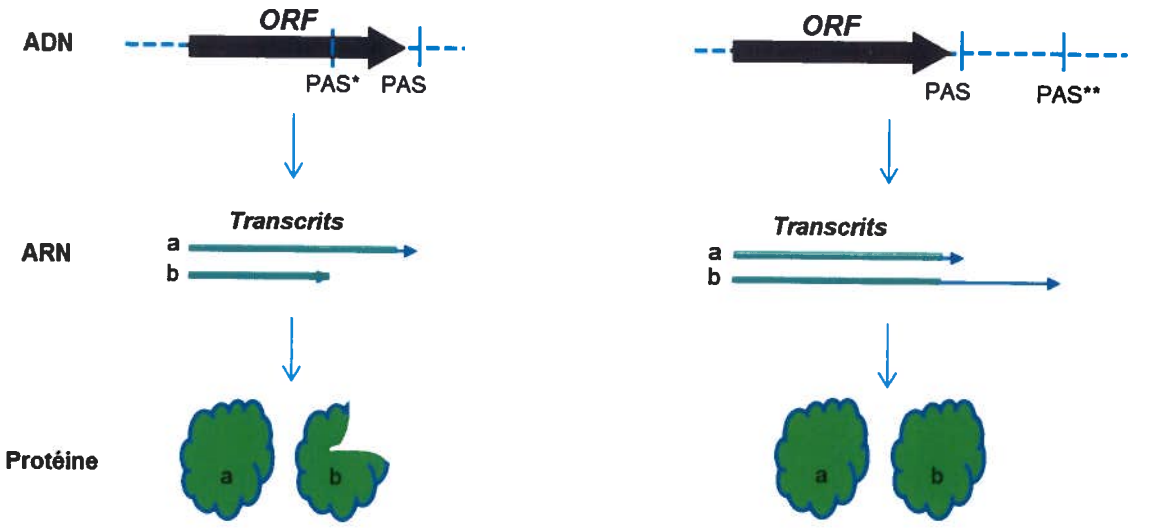

Fig. A1. Schéma représentatif de la polyadénylation alternative. Le panneau de gauche montre la conséquence de la polyadénylation alternative où l'un des PAS est dans l'ORF (flèche noire). « PAS » indique le site de polyadénylation qui est juste après l'ORF, et « PAS* » indique le site de polyadénylation présent dans l'ORF. Le panneau de droite montre la conséquence de la polyadénylation alternative dans le 3'UTR, où « PAS** » indique le site de polyadénylation utilisé en aval du premier PAS du gène. Les transcrits générés par ces deux processus sont représentés par les flèches bleues ci-dessous avec les rectangles verts soulignant l'ORF. Les protéines générées sont représentées par les nuages verts dans les parties inférieures des panneaux.

La machinerie de la polyadénylation

La machinerie de la polyadénylation comprend les éléments en *cis* d'ARNm ainsi que six facteurs de polyadénylation. Les éléments en *cis* comprennent la séquence hexamèrique AAUAAA, le site de clivage, l'élément riche en G/U, ainsi que des éléments riches en U et en G qui sont moins bien conservés (Hu et al., 2005; Lutz, 2008; revue de Mandel, Bai, and Tong, 2008; revue de Takagaki and Manley, 1997; Zhao, Hyman, and Moore, 1999).

Les facteurs de polyadénylation comprennent : le facteur de spécificité de clivage et de polyadénylation (CPSF), le facteur de stimulation de clivage (CstF), les facteurs de clivage I et II (CFI_m et $CFII_m$ respectivement), la polymérase polyA (PAP), le facteur de liaison de la queue polyA et symplekine (Fig. A2). CPSF contient la sous-unité CPSF-160 qui reconnait la séquence hexamèrique, la sous-unité CPSF-73 qui semble être l'endonucléase responsable pour le clivage (revue de Dominski, 2007), ainsi que la sous-unité CPSF-100, une protéine d'échafaudage du complexe (Kyburz et al., 2003). CstF contient la sous-unité CstF-64 qui a un domaine de liaison d'ARN et qui lie la région G/U riche des transcrits (revue de Takagaki and Manley, 1997). Le facteur de clivage CFI_m contient quatre sous-unités (CFI_m -25, -59, -68 et -72), parmi eux, la sous-unité CFI_m-25 interagit avec PAP et PABP, et lie l'ARN à l'aide de CFI_m-68 (Kim and Lee, 2001) (Dettwiler et al., 2004). Le facteur de clivage $CFII_m$ contient deux sous-unités (hClp1 et hPcf11). hClp1 qui aide à lier $CFII_m$ et CPSF, et semble aussi avoir un rôle dans le clivage des transcrits (de Vries et al., 2000; Hollingworth et al., 2006). La polymérase PAP est responsable pour l'ajout de la queue polyA. La PABP est nécessaire pour le clivage et la polyadénylation et contient quatre motifs de liaison à l'ARN (Deo et al., 1999; Meyer, Urbanke, and Wahle, 2002). Et finalement, simplekine semble être une protéine d'échafaudage qui interagit avec CPSF et CstF (Takagaki and Manley, 2000). Tous ces facteurs forment un système efficace et versatile pour le traitement et la maturation de l'ARN.

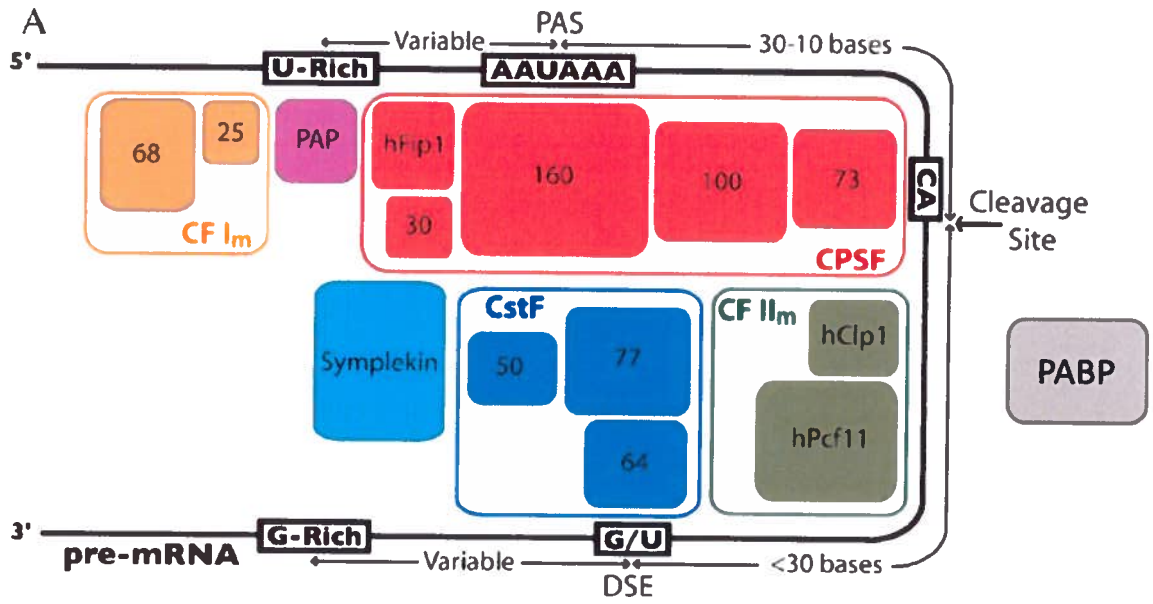

Fig. A2. Diagramme représentatif de la machinerie de polyadénylation. La ligne noire représente l'extrémité 3' du pré-ARNm et les différents éléments agissant en cis. Les facteurs de polyadénylation différents et leurs sous-unités sont également représentés comme suit: CPSF (rouge), CstF (bleu), CFI_m (jaune) et $CFII_m$ (vert), le PAP (pourpre), le PABP (gris) et symplekine (bleu clair). Les nombres dans les rectangles font référence aux noms des sous-unités et correspondent aussi à leurs masses moléculaires (tiré de Mandel, Bai, and Tong, 2008).

La polyadénylation alternative chez le VHS-1

La polyadenylation alternative est utilisée pour réguler l'expression des gènes dans plusieurs systèmes viraux incluant le cytomégalovirus (Goins and Stinski, 1986; Stamminger, Puchtler, and Fleckenstein, 1991), l'adénovirus (Mann, Weiss, and Nevins, 1993; Nevins and Wilson, 1981), le virus du polyome (Hyde-DeRuyscher and Carmichael, 1990) ainsi que le virus du papillome (Kennedy, Haddow, and Clements, 1990).

La polyadénylation alternative a été observée chez plusieurs gènes du VHS-1 incluant, *UL2* (Singh and Wagner, 1993), *UL24* (Cook and Coen, 1996), *UL38* (Anderson et al., 1981) et *UL52* (Uprichard and Knipe, 1996).

Malgré que ce n'est pas relié directement à la polyadénylation alternative chez le VHS-1, il a été démontré que des protéines nucléaires se lient plus efficacement l'ARNm viral durant l'infection (McGregor et al., 1996). De plus, il a été rapporté qu'il y a une compétition cinétique entre les facteurs de polyadénylation et les facteurs d'épissage. Les facteurs d'épissage sont redistribués durant l'infection et ne sont plus présents aux sites de transcription, ceci permet un meilleur accès aux facteurs de polyadénylation dans le cas où ces deux types de facteurs étaient en compétition (Martin et al., 1987). Ce processus semble être dépendant d'ICP27. En effet, il a été démontré qu'ICP27 stimule le traitement de certain PAS, comme c'est le cas pour *UL38* et *UL44*, ayant des PAS faibles (McGregor et al., 1996). Par contre, une mutation inhibant l'expression d' ICP27 n'avait aucun effet sur les transcrits courts d'*UL24*, qui utilisent eux aussi un PAS faible, et même semblent favoriser l'utilisation du PAS faible d'*UL24* (Hann et al., 1998).

PROBLÉMATIQUE

Le VHS-1 est un virus a ADNdb qui code pour plus de 80 gènes et qui nécessite donc une régulation génétique efficace. Un des modes de régulation utilisé par le VHS-1 est celui de la polyadénylation alternative: un évènement post-transcriptionnel qui permet à un gène d'utiliser deux PAS. Ce processus génère des transcrits qui diffèrent dans leurs 3'UTR, ce qui peut affecter l'efficacité de la traduction entre autres. Un des gènes qui est régulé par la polyadenylation alternative est le gène *UL24*. Les transcrits courts d'*UL24* sont générés à partir du PAS signal d'*UL24* dans les temps IE, tandis que les transcrits longs et polycistroniques sont générés à partir du PAS d'*UL26* dans les temps L. Similairement, *UL38* a aussi deux types de transcrits qui sont générés, ou bien à partir de son propre PAS, ou bien à partir de celui d'*UL39/UL40*.

Notre hypothèse est que (I) la transition entre l'expression des transcrits courts et des transcrits long est due à un changement dans le niveau basal de l'expression des facteurs de polyadenylation. Aussi, (II) que le niveau d'association des facteurs de polyadenylation aux transcrits viraux diffère durant l'infection et d'un PAS à un autre.

Pour vérifier ces hypothèses, nous avons testé des cellules Vero et HeLa mock-infectées ou infectées pendant 3, 6, 9, 12 et 16 hrs pour les niveaux d'expression de divers facteurs de polyadenylation, par immunobuvardage de type Western. De plus, un système de purification par affinité de protéines se liant à l'ARN a été développé pour évaluer les protéines qui se lient aux régions 3'UTR des différents transcrits.

METHODES ET RESULTATS

Évaluation des niveaux d'expression basals de facteurs de polyadénylation durant l'infection par le VHS-1

Préparation des lysats cellulaires

Les transcrits courts d'*UL24* sont exprimés avec une cinétique précoce et les transcrits longs avec une cinétique tardive. Donc, pour mieux comprendre la polyadénylation alternative des transcrits d'*UL24*, nous avons évalué les niveaux d'expression basale de différentes sous-unités de la machinerie de polyadénylation.

Pour ce faire, nous avons récolté des lysats de cellules Vero et HeLa à partir de pétris de 60 mm en utilisant le tampon RIPA (avec 500 mM NaCl). Les lysats étaient récoltés de cellules non-infectées, ainsi qu'infectés avec un MOI de 5, pour 3, 6, 9, 12 et 16 hrs. Ceci nous a permis d'évaluer les niveaux des facteurs cellulaires de polyadénylation durant toute la cascade temporelle de l'infection.

Immunobuvardage de type Western pour les facteurs de polyadénylation

Les lysats cellulaires obtenus ont été testés afin d'évaluer les niveaux des facteurs de polyadénylation et ceci à l'aide des anticorps disponibles commercialement. Plusieurs anticorps testés présentaient des marquages peu ou pas-spécifiques. Par contre, nous avons pu évaluer une sous-unité pour chacun des trois facteurs CPSF, CstF et CFI_m (Fig. A3-A5). Les membranes ont aussi été testées pour la Ɣ-tubuline, comme témoin. Les résultats présentés sont représentatifs de ceux obtenus dans au moins trois différentes expériences répétées.

Malgré certaines fluctuations dans l'intensité des bandes dans nos immunobuvardages, notamment entre les lysats « mock » et les lysats à 3 hpi, nos résultats suggèrent qu'il y a aucun changement reproductible dans le niveau d'expression basal des sous-unités testées jusqu'à 12 hpi dans les deux lignées cellulaires testées (Fig. 11-13). Par contre, il semble y avoir une légère baisse du niveau d'expression de CPSF-100 dans les cellules Vero à 16 hpi, mais cette baisse se manifeste à un temps tardif comparé au temps où l'on observe une diminution dans

l'expression des transcrits courts (10 hpi). Donc, la baisse observée pour le facteur CPSF-100 ne peut être la cause pour la diminution de l'usage du PAS d'*UL24*.

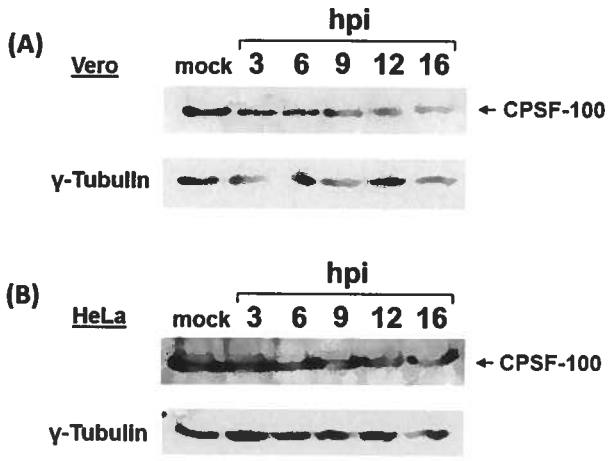

Fig. A3. Les tests d'immunobuvardage de type Western Blot montrant les niveaux du facteur CPSF-100 au cours de l'infection par VHS-1. Analyse des lysats des cellules Vero (A) et des cellules HeLa (B) qui sont soit non-infectées, ou infectées aux différents temps indiqués. Panneaux supérieurs montrent des niveaux du facteur CFSF-100 et les panneaux inférieurs représentent les résultats pour le contrôle de charge, la γ-tubuline.

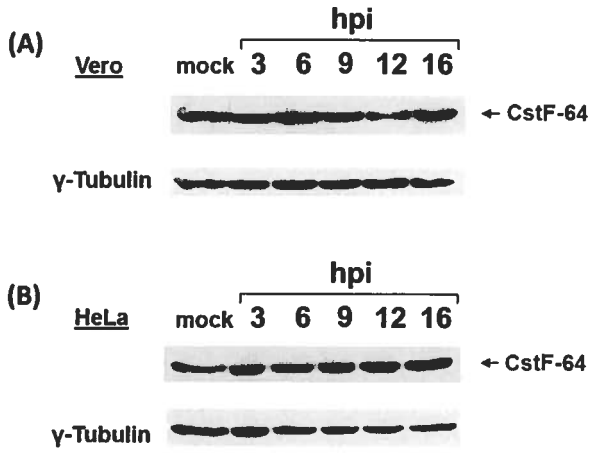

Fig. A4. Les tests d'immunobuvardage de type Western Blot montrant les niveaux du facteur CstF-64 au cours de l'infection par VHS-1. Analyse des lysats des cellules Vero (A) et des cellules HeLa (B) qui sont soit non-infectées, ou infectées aux différents temps indiqués. Panneaux supérieurs montrent des niveaux du facteur CstF-64 et les panneaux inférieurs représentent les résultats pour le contrôle de charge, la γ-tubuline.

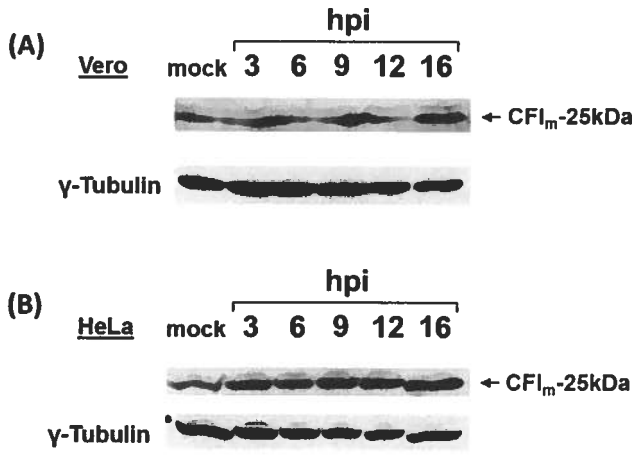

Fig. A5. Les tests d'immunobuvardage de type Western Blot montrant les niveaux du facteur CFI_m-25 au cours de l'infection par VHS-1. Analyse des lysats des cellules Vero (A) et des cellules HeLa (B) qui sont soit non-infectées, ou infectées aux différents temps indiqués. Panneaux supérieurs montrent des niveaux du facteur CFI_m-25 et les panneaux inférieurs représentent les résultats pour le contrôle de charge, la γ-tubuline.

En résumé, la polyadénylation alternative menant à l'expression des transcrits courts d'*UL24* aux temps précoces durant l'infection et l'expression des transcrits long tard durant l'infection, ne peut être attribuée à un changement dans les niveaux basales d'expression des facteurs CPSF-100, CstF-64 et CFI_m-25.

Purification par affinité avec streptavidine de protéines se liants à l'ARN

Un système de purification par affinité a été utilisé pour isoler des facteurs qui interagissent avec les régions contenant les PAS des transcrits viraux affectés par la polyadenylation alternative.

Description de l'aptamère

La sélection des aptamères *in vitro* a permis la découverte de séquences d'ARN et d'ADN, qui peuvent lier efficacement des petites molécules ou des macromolécules (Wilson and Szostak, 1999). Srisawat et Engelke ont décrit des aptamères pouvant lier spécifiquement la streptavidine et le Sephadex (Srisawat and Engelke, 2001; Srisawat and Engelke, 2002; Srisawat, Goldstein, and Engelke, 2001). Nous avons choisi le système de purification de la streptavidine due à sa forte affinité pour l'aptamère S1 (70nM) (Fig. A6). De plus, ce système permet une élution spécifique et relativement plus facile du complexe ARN-protéines dans des conditions non-dénaturantes que permet le système Sephadex. La liaison non-spécifique de la biotine, présente dans le lysat, à la streptavidine peut être contrôlée en saturant la biotine avec l'avidine du blanc d'œufs.

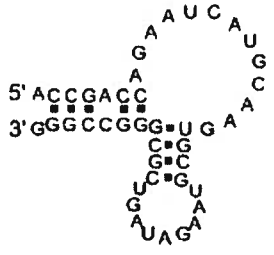

Fig. A6. L'aptamère de streptavidine. La séquence nucléotidique et la structure 2-D de l'aptamère de la streptavidine S1.

Construction du plasmide

Pour pouvoir étudier la polyadénylation alternative des gènes viraux *UL24*, *UL26*, *UL38* and *UL40*, les régions contenant le PAS de chacun des quatre gènes ont été clonées dans le plasmide pBluescript SK+ (Stratagene) (Fig. A7 et A8). Des plasmides avec la séquence d'ADN inversée (d'*UL24* et d'*UL40*) ont aussi été construits pour être utilisés comme témoins négatif. Les séquences des régions contenant le PAS des divers gènes sont présentées dans la figure A8. Chaque plasmide contenait aussi la séquence de 44 nt de l'aptamère reconnu par la streptavidine. Les plasmides générés ont été nommés pSK-pAUL24-tag, pSK-pAUL26-tag, pSK-pAUL38-tag, pSK-pAUL40-tag, pSK-pAUL24inv-tag et pSK-pAUL40inv-tag. Ces plasmides ont servi comme matrice pour la synthèse d'ARN *in vitro*.

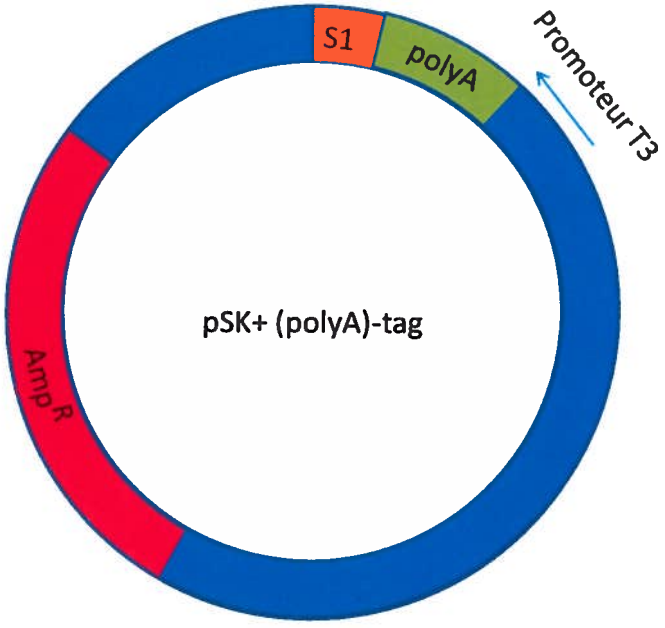

Fig. A7. Carte des constructions plasmidiques de polyA-S1 tag. pBluescript SK + a été utilisé comme vecteur de départ, il contient le gène de résistance à l'ampicilline. Les fragments contenant le PAS (vert) des différents gènes ont été insérés dans le site de clonage multiple de pSK +, en accord avec le promoteur T3, sauf pour les contrôles négatifs qui ont été insérés dans l'orientation inverse. La séquence codant pour le tag S1 (orange) a été clonée dans le site de restriction Xho I à proximité.

***UL24* PAS**

5' UGC CAA CGC CAG ACG CCG GUC CGC UGU GCC AUC GNU CCC CUU CAU CCC ACC CCC AUC UUG UCC CCA AAU AAA ACA AGG UCU GGU AGU UAG GAC AAC GAC CGC AGU UCU CGU GUG UUA UUG UCG CUC UCC GCC UCU CGC AGA 3'

***UL26* PAS**

5' CGA UUU GUU CGU CUC UCA GAU GAU GGG GGC CCG CUG AUU CGC CCC GGU CUU UGG UAC CAU GGG AUG UCU UAC UGU AUA UCU UUU UAA AUA AAC CAG GUA AUA CCA AAU AAG ACC CAU UGG UGU AUG UUC UUU UUU UUU UUU UAU UGG GAG GGG CGG GUA GGC GGG UAG CUU UAC AAU GCA AAA GCC UUU GAC GUG GAG GAA GG 3'

***UL38* PAS**

5' UGA GCG UAG CAA ACG CCC CGC CCA CAC AAC GCU CCG CCC CCA ACC CCU UCC CCG CUG UCA CUC GUU GUU CGU UGA CCC GGG CGU CCG CCA AAU AAA GCC ACU GAA ACC CGA AAC GCG AGU GUU GUA ACG UCC UUU GGG CGG GAG GAA GCC ACA AAA UGC 3'

***UL40* PAS**

5' AGU GUC GCA GCA CCU CCU ACG CCG GGG CGG UCG UCA ACG AUC UGU GAG *G*GU CGC GGC GCG CUU CUA CCC GUG UUU GCC CAU AAU AAA CCU CUG AAC CAA ACU UUG GGU CUC AUU GUG AUU CUU GUC AGG GAC G 3'

- Region U-riche
- PAS
- Region GU-riche
- Region G-riche

Fig. A8. Séquences contenant les PAS des gènes viraux. Les divers séquences clonées dans le vecteur pBluescript SK+ sont indiquées. Les séquences soulignées montrent l'emplacement des éléments agissant en cis.

Préparation de la matrice d'affinité

Les étapes de la purification par affinité sont illustrées dans la figure A10. Les transcrits qui seraient utilisés sur les colonnes d'affinité ont été générés à l'aide du kit T3 MEGAscript™ (Ambion). L'ARN synthétisé contenait la région 3'UTR ayant le PAS avec l'aptamère S1 fusionné à la région 3'. L'ARN a été synthétisé à partir de la matrice d'ADN linéarisé à l'aide d'une digestion avec une enzyme de restriction. La quantité d'ARN obtenue par réaction était de 100-250 ug. Des aliquots d'ARN synthétisé ont été évalués en migrant sur un gel de polyacrylamide dénaturant pour confirmer la taille de l'ARN. Par contre, nous avons aussi observé des bandes de plus grandes tailles, potentiellement causées par des structures d'ARN non-dénaturés, ainsi que des bandes de plus petites tailles, possiblement due à des coupures dans la matrice d'ADN utilisée (résultats non montrés). Divers tampons ont été testés pour leur compatibilité avec le système de purification base sur l'aptamère S1. Nous avons trouvé que seul le tampon suggéré par les auteurs Srisawat et Engelke a permis la liaison de l'aptamère aux billes (Srisawat and Engelke, 2001).

Préparation des lysats

Des lysats ont été récoltés de cellules non-infectées ainsi que de cellules infectées à une MOI de 5 pour 5 ou 10 hrs. Les lysats ont été préparés avec le kit « complete M-lysis » de Roche. Des lysats provenant de $1x10^8$ cellules ont été utilisés pour chaque matrice. Comme le tampon de lyse du kit n'était pas compatible avec notre stratégie de purification par affinité, une dialyse a été effectuée pour échanger le tampon avec celui qui a été recommandé pour l'aptamère S1 de la streptavidine. Une portion des protéines s'est précipitée à cette étape, comme a été observé visuellement à la fin de la dialyse. Ces précipités ont été enlevés par centrifugation à 12,000 rpm pour 30 min (JA-20 rotor).

Purification par affinité de protéines se liant à l'ARN

Dans le contexte de ce mémoire, seulement les PAS d'*UL24* et d'*UL26* ont été testés.

Nous voulions confirmer que l'ARN synthétisé *in vitro* ayant l'aptamère S1 était bien transcrit et capable de lier la streptavidine. En utilisant les mêmes conditions de liaison pour la purification par affinité, nous avons incubé les billes avec l'ARN, dans l'absence du lysat. Après deux lavages, une élution de 20 min avec la biotine (5 mM) et une deuxième élution de 5 min avec l'urée (8 M) ont été effectuées. Les éluats ont été migrés par électrophorèse sur gel acrylamide dénaturant, et l'ARN a été coloré à l'aide de bromure d'éthidium. Nous avons obtenu une élution spécifique de l'ARN avec la biotine (fig. A9-EB). Une bande plus intense a été observée

avec l'élution de l'urée (fig. A9-EU); suggérant qu'une certaine portion de l'ARN était encore liée après l'élution relativement courte de 20 min avec la biotine.

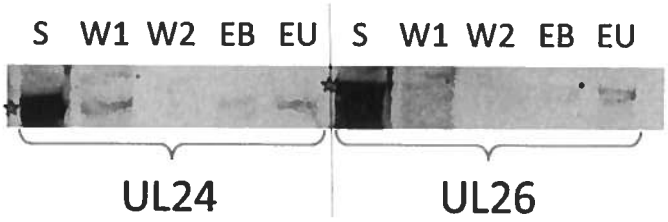

Fig A9. Gel ARN de la purification par affinité avec ligand seulement. À gauche, résultats pour l'ARN pAUL24-tag, et à droite, résultats pour l'ARN pAUL26-tag. S, W1, W2, EB et EU, signifient surnageant après l'addition de l'ARN, premier lavage, deuxième lavage, élution avec biotine et élution avec l'urée, respectivement. Les étoiles démontrent les niveaux de migration attendus des transcrits.

Pour la purification à grande échelle (Fig. A10), les billes de streptavidine ont été équilibrées dans le tampon de liaison et par la suite, incubées pendant 1 hr en présence des transcrits générés *in vitro* contenant les PAS d'*UL24*, d'*UL26* ou leur contrôle négatif UL24inv, fusionnés à l'aptamère S1. Ensuite, les billes ont été lavées et les lysats cellulaires préparés ont été ajoutés. Le tout a été incubé pendant 1 hr et suite à des lavages avec le tampon de liaison, deux élutions ont été effectuées; une première avec la biotine (5mM) et une deuxième avec l'urée (8M).

Les protéines présentes dans les éluats ont été migrées sur un gel d'acrylamide 10%, dénaturant avec SDS, et colorées à l'argent (Figs. A11 et A12). Des bandes spécifiques ont été détectées suite à l'élution avec la biotine. Dans les deux élutions, les bandes de protéines étaient plus intenses dans les échantillons de 10 hpi ; ceci est probablement dû à la présence de plus de protéines dans ces lysats par une erreur de manipulation. Il semblait aussi y avoir plus de protéines dans les éluats d'ARN pAUL26-tag que de ceux correspondant à pAUL24-tag. Les bandes les plus intenses dans les échantillons pAUL24 et pAUL26 comparées aux échantillons pAUL24inv ont été sélectionnées pour analyse par chromatographie liquide et spectrométrie de masse (LC-MS) (Figs. A11 et A12).

Purification par affinité de protéines liants l'ARN

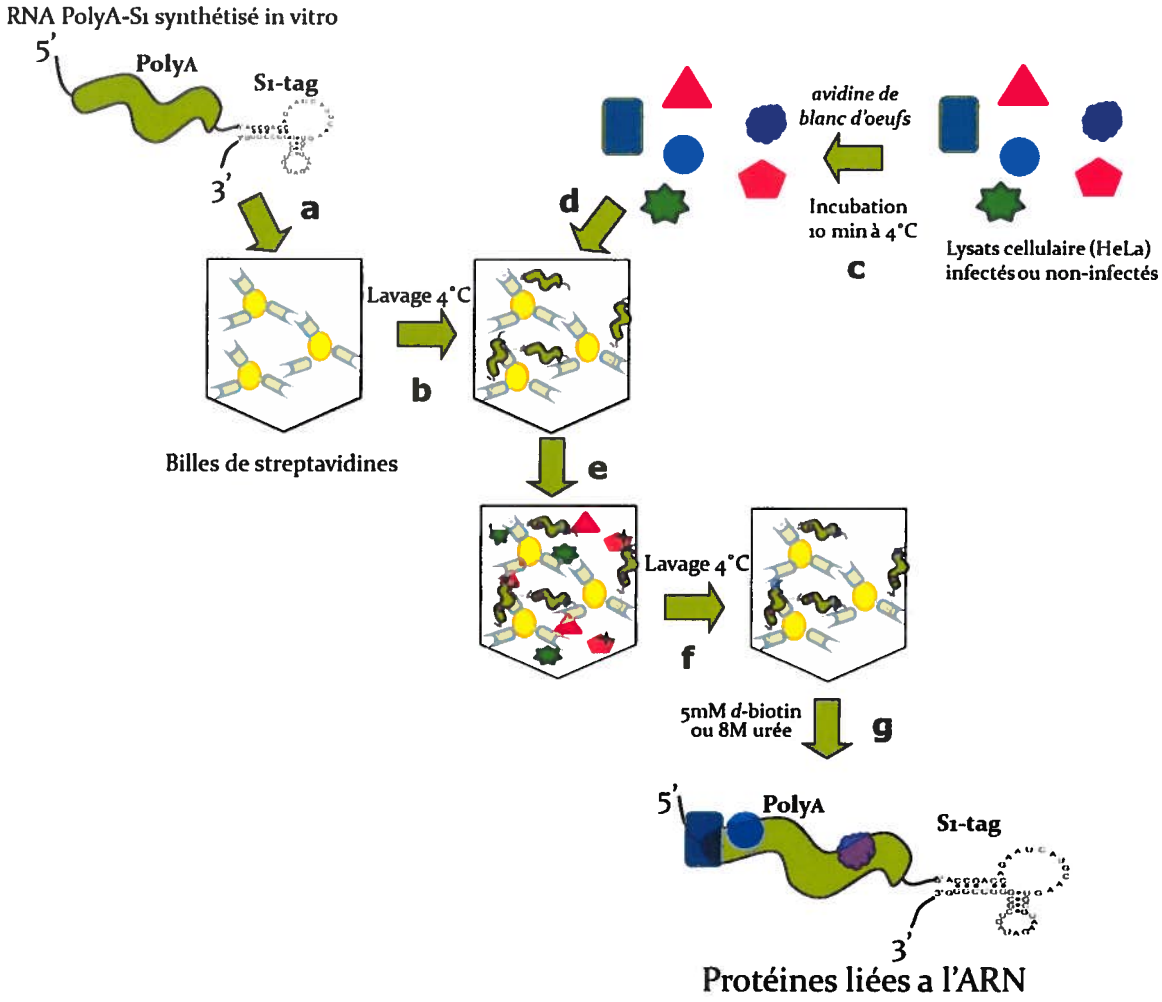

Fig. A10. Schéma représentatif du système de purification par affinité. Les étapes de purification d'affinité: (a) l'addition d'ARN transcrit à des billes de streptavidine équilibrées et son incubation, (b) des lavages pour éliminer l'excès d'ARN non lié, (c) les lysats des cellules HeLa pré-incubés avec l'avidine de blanc d'œuf, (d) l'addition des lysats des cellules HeLa aux billes liées à l'ARN (e) l'incubation, (f) les lavages pour éliminer les protéines non liées, et (g) les élutions pour obtenir les complexes de protéines qui se liées à l'ARN.

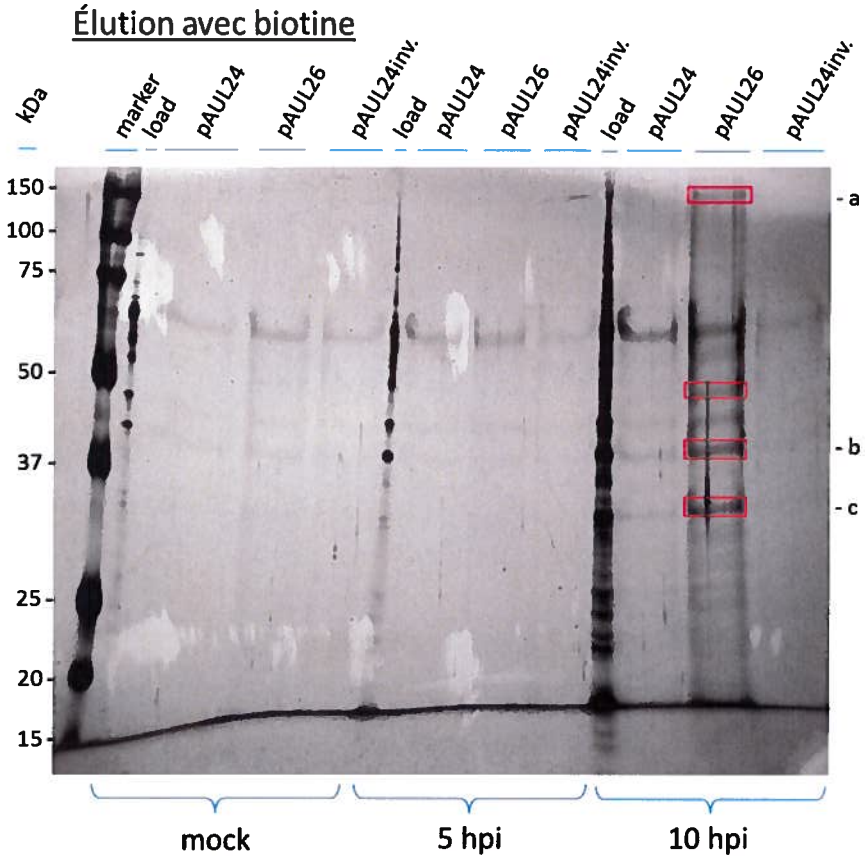

Fig. A11. L'analyse des gels colorés au nitrate d'argent montrant les éluats de biotine. Le gel montre les éluats de cellules non-infectées (mock) et les cellules infectées pendant 5 et 10 hpi. Pour chaque temps d'infection indiqué, l'expérience a été réalisée avec des matrices d'ARN correspondant à pAUL24-tag, pAUL26-tag et le contrôle négatif pAUL24inv-tag. Les bandes ainsi extraites du gel sont mises en évidence par les rectangles rouges. Chaque bande est représentée par une lettre utilisée comme nom d'échantillon lors de l'analyse par LC-MS. Un aliquot de lysat analyse est présent dans la voie « load ». Les bandes du marqueur de poids moléculaire « marker » sont indiquées à gauche du gel.

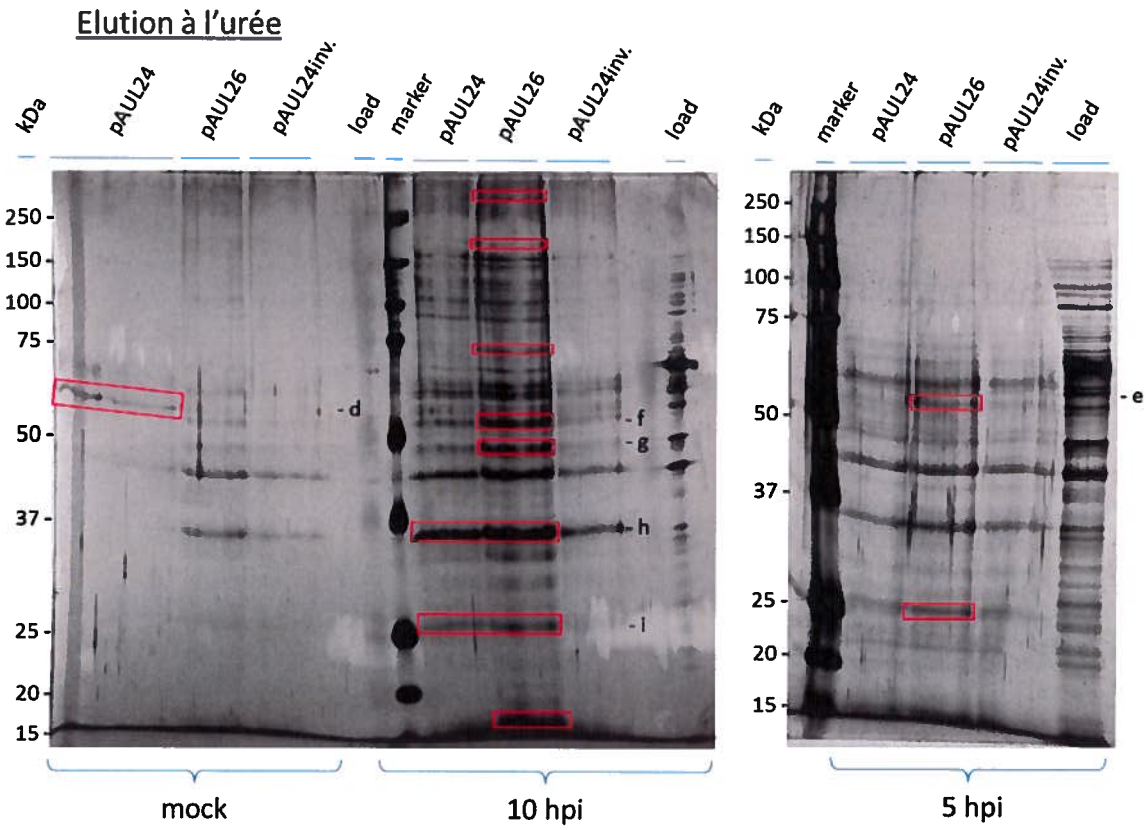

Fig. A12. L'analyse des gels colorés au nitrate d'argent montrant les éluats de l'urée. Le gel montre les éluats de cellules non-infectées (mock) et les cellules infectées pendant 5 et 10 hpi. Pour chaque temps d'infection indiqué, l'expérience a été réalisée avec des matrices d'ARN correspondant à pAUL24-tag, pAUL26-tag et le contrôle négatif pAUL24inv-tag. Les bandes ainsi extraites du gel sont mises en évidence par les rectangles rouges. Chaque bande est représentée par une lettre utilisée comme nom d'échantillon lors de l'analyse par LC-MS. Un aliquot de lysat analyse est présent dans la voie « load ». Les bandes du marqueur de poids moléculaire « marker » sont indiquées à gauche du gel.

Analyse par LC-MS

Les analyses par LC-MS ont été effectuées par le service de spectrométrie de masse de l'INRS-Institut-Frappier. Les bandes découpées pour analyse ont été digérées par la trypsine. Par la suite, les peptides obtenus ont été séparés par chromatographie liquide en phase inverse (Agilent Nanopump system) et identifiés par spectrométrie de masse (hybrid linear ion trap-triple quadrupole mass spectrometer). Pour l'analyse, un score de MS de 43 ou plus a été considéré comme significatif. Parmi les 15 bandes découpées, huit ont permis l'identification de protéines, notamment l'actine, la chaperonine, la dynèine, l'enolase, l'HSP-27, la kinase « leucine-rich repeat serine/threonine-protein» 2 (LRRK2), la myosine, le facteur nucléaire 45 (NF45), la transglutaminase, le facteur d'élongation de traduction 1 (eEF1A), la tubuline, ainsi que des protéines inconnues (Tableau A1). De plus, nous avons inclus dans le tableau le facteur d'élongation de l'ARN (ELL) malgré que son score est un peu moins que 43. De plus, nous avons identifié des isoformes du contaminant kératine, non-inclus dans le tableau.

Certaines des protéines qui sont abondantes dans la cellule ont été éliminées comme contaminants qui se lient non-spécifiquement à la colonne : l'actine, la tubuline, la myosine, la dynèine, l'enolase, la transglutaminase et l'HSP-27. Étonnement, aucun facteur de polyadénylation n'a été identifié malgré que la présence de CPSF-100 et CstF-64 a été confirmée par immunobuvardage de type Western (résultats non-inclus). Les protéines ELL, NF45 et eEF1A sont particulièrement intéressantes car elles ont des propriétés de liaison d'acides nucléiques.

Bande	Élution	Protéines	Nombre de peptides	Score de la protéine	Taille de la bande (kDa)	Taille prédite de la protéine (kDa)
A	Biotine/ UL26	elongation factor RNA polymerase II	2	40	~125	80
B	Biotine/ UL26	mutant beta-actin (beta'-actin) alpha-actin gamma-actin actin, beta-like 2 *similar to* actin, gamma 1 actin-like protein actin-like protein	4 4 3 3 1 6 2	208 187 156 150 86 63 46	~40	42
C	Biotine/ UL26	Unnamed	3	46	~32	-
D	Urée/ UL24	hCG2003116, isoform CRA_a	3	49	~57	-
E	Urée/ UL26	Unnamed	3	51	~ 60	-
		Leucine-rich repeat serine/threonine-protein kinase 2	4	50		286
		KIAA1170	3	45		-
F	Urée/ UL26	tubulin, beta 5 tubulin, beta, 2 Unnamed protein (tubulin motifs) alpha-tubulin tubulin, alpha 1B hCG27371-tubulin hCG1747788-tubulin	9 7 7 7 7 2 1	637 507 501 345 343 87 57	~52	48-57
		chaperonin containing TCP1 subunit 2	1	53		60
G	Urée/ UL26	2-phosphopyruvate-hydratase alpha-enolase; carbonate dehydratase Unnamed protein product (enolase) neurone-specific enolase	3 1 1	202 97 45	~47	48
		glucocorticoid receptor AF-1 specific elongation factor *similar to* eukaryotic translation elongation factor 1 alpha 1 isoform 4	3 3	146 118		50
H	Urée/ UL24 & UL26	Actin, gamma 1 propeptide Gamma-actin Actin, beta-like 2 Actin-like protein *similar to* actin, gamma 1	9 7 5 2 1	363 264 151 92 50	~35	42
		hCG2018327	3	52		-
		KIAA0799 (myosin motif identified)	5	52		variable
		NF45 protein	1	47		45
		dynein, axonemal, heavy chain 3	3	43		530
I	Urée/ UL24 & UL26	heat shock protein 27	2	87	~26	27
		prostate-specific transglutaminase 4	3	50		77

Tableau A1. Résumé de l'analyse par LC-MS des produits de la purification par affinité. Liste des protéines identifiées des éluats de la biotine ou de l'urée des matrices d'affinité d'ARN.

DISCUSSION ET CONCLUSIONS

L'impact de l'infection par le VHS-1 sur les niveaux d'expression basale de facteurs de polyadénylation

Nous voulions tester notre hypothèse que la transition entre l'expression des transcrits courts aux transcrits longs d'UL24 due à la polyadenylation alternative est due à des changements dans les niveaux d'expression basale de composants de la machinerie de clivage et de polyadénylation. Nos études d'immunobuvardages de type Western effectuées sur les facteurs CPSF-100, CstF-64 et CFI_m-25 ont démontré qu'il n'y aucun changement reproductible dans le niveau d'expression basale de ces facteurs durant les premières 12 heures d'infection. La transition entre l'utilisation du PAS faible d'*UL24* au signal plus fort d'*UL26* se fait avant 10 hrs post-infection, donc les niveaux d'expression de ces facteurs ne peuvent pas être la cause de cette transition. Par contre, il y a d'autres facteurs qui n'ont pas pu être évalués, car les anticorps ne permettaient pas d'avoir un signal spécifique ou ils n'étaient pas disponibles commercialement. Nous ne pouvons donc pas exclure la possibilité qu'un changement dans leur niveau d'expression soit impliqué dans ce processus. D'autres anticorps plus spécifiques devraient être utilisés pour tester cette hypothèse.

De plus, même si les niveaux basaux d'expression des facteurs ne changent pas durant l'infection, il est possible que le niveau d'association avec d'autres facteurs ou avec les transcrits soit affecté par des changements post-traductionnels durant l'infection. En effet, chaque facteur du complexe de polyadénylation est affecté par ce type de modification (Millevoi and Vagner, 2010).

Un autre mécanisme pouvant mener à un changement dans le traitement de PAS est celui de la relocalisation de certain facteurs de polyadénylation les rendant inaccessibles au site de polyadénylation, comme se fait chez les cytomégalovirus humains, un membre des betaherpesvirinae (Gaddy et al., 2010). Par ailleurs, chez le VHS-1, la protéine PABP cytoplasmique est relocalisée durant l'infection et s'accumule au noyau (Dobrikova et al., 2010). Ceci est une manière pour réguler transcriptionnellement et post-transcriptionnellement le traitement et l'export nucléaire ARNm.

La purification par affinité des protéines liant l'ARN utilisant l'étiquette de streptavidine

Pour évaluer le niveau d'association des facteurs de polyadénylation à différents PAS viraux et à différents temps post-infection, un système de purification par affinité a été utilisé pour identifier les protéines qui interagissent avec les régions 3'UTR des transcrits courts et longs d'*UL24*. Les résultats attendues étaient que des protéines qui interviennent dans la stimulation de la polyadénylation se lieraient de façon plus importante à l'ARN 3' non-traduite d'*UL26* que celui d'*UL24*, puisque ce dernier représente un PAS plus faible. De même, nous nous attendions à possiblement détecter des protéines inhibitrices dans les éluats du 3'UTR d'*UL26* à des temps précoce ou dans les éluats du 3'UTR d'*UL24*, puisque dans ces cas, les PAS ne fonctionnent pas de façon très efficace.

La stratégie employée a permis l'isolement des protéines associées aux ARNs viraux. Cette expérience représente donc la preuve de concept de notre stratégie, et nous avons pu identifier des protéines avec des propriétés de liaison à l'acide nucléique, entres autres. Par contre, la majorité des bandes envoyées pour analyse pas LC-MS étaient faibles et ils semblaient être à la limite de détection de l'instrument.

Le système d'affinité utilise peut encore être optimisé à différents niveaux. Par exemple, au niveau du choix de la région 3'UTR à évaluer pour les différents gènes afin de s'assurer de bien avoir tous les éléments nécessaires pour le traitement efficace en 3'. De plus, le tampon de lyse commercial utilisé avait ses limitations et nécessitait une dialyse qui a causé une perte significative de protéines du lysat. Donc une stratégie alternative de lyse pourrait être évaluée telle qu'une lyse par sonication. Ceci permettra d'obtenir de meilleurs résultats par ce système de purification.

Identification des protéines purifiées par affinité à l'aide de LC-MS

Nous avons identifié plusieurs protéines qui lient des acides nucléiques à partir de notre système de purification par affinité de complexes ARN-protéines.

eEF1A

eEF1A a été identifié avec un score de MS relativement élevé de 118, et a été excisé du puits pAUL26 à 10 hpi. eEF1A est une protéine multifonctionnelle qui joue un rôle dans la traduction et dans l'export nucléaire de tRNAs matures aux ribosomes (Grosshans, Hurt, and Simos, 2000; Negrutskii and El'skaya, 1998; Reed, Wastney, and Yang, 1994). Aussi, elle est impliquée dans l'export nucléaire de la protéine PABP1 de façon dépendante sur l'ARN pol II (Khacho et al., 2008). De plus, cette protéine joue un rôle important dans la fidélité de la traduction (Carr-Schmid et al., 1999), et comme durant l'infection le niveau de traduction de protéines virales est haut, il est possible que cette protéine soit nécessaire pour assurer la fidélité de la traduction.

Cette bande était présente dans les trois puits à 10 hpi, mais était plus intense chez les échantillons polyA. Ceci confirme que l'on peut isoler des complexes ARN-protéines, mais qu'il y a encore des optimisations à faire pour isoler des facteurs qui lient spécifiquement le PAS.

NF45

NF45 est une autre protéine qui a été identifiée et qui lie l'acide nucléique (score 47). Elle a été isolée des puits pA-UL24 et pA-UL26 à 10 hpi et les bandes des deux échantillons ont été combinées pour analyse par LC-MS. NF45 est un activateur de transcription hautement conservé qui forme un complexe avec NF90 et NF110 (Parrott et al., 2005; Zhao et al., 2005). Ces trois composantes contiennent un « dimerisation zinc finger nucleic acid-binding domain » et lient l'ARN double-brin (ARNdb) (Meagher et al., 1999). Son partenaire NF90 interagit avec l'ARNdb ainsi que l'ARN simple-brin structuré (Liao, Kobayashi, and Mathews, 1998; Parrott et al., 2005; Reichman and Mathews, 2003). NF90 lie aussi les 3'UTR et aide à stabiliser ARNm (Shi et al., 2005; Shim et al., 2002). Étant donné que NF90 et NF45 forment des complexes, ceci suggère un rôle pour NF45 dans la régulation post-transcriptionnelle et confirme encore une fois que ce système peut être utilisé pour isoler des complexes ARN-protéines.

ELL

ELL a aussi été identifié par MS à partir du puits pA-UL26. Malgré que son score (40) était un peu plus bas que 43, et que son niveau de migration sur gel ne correspondait pas à sa taille prédite, nous l'avons inclus dans l'analyse comme exemple de facteur qu'on s'attendait à identifier dans cette étude. ELL régule l'activité de ARN pol II durant la phase d'élongation de transcription (Shilatifard et al., 1997a). Il stimule l'élongation en supprimant le « transcriptional pausing ». Ceci pourrait faire en sorte que la transcription continue après le premier PAS d'*UL24* et s'arrête seulement au PAS d'*UL26*. Malheureusement, nous sommes à la limite de détection pour l'analyse des échantillons et donc des améliorations à la technique seraient nécessaires pour identifier avec plus de certitude les protéines impliquées dans la polyadénylation alternative.

Pour confirmer ces résultats, en premier lieu les purifications à grande échelle devront être répétées. Par ailleurs, les résultats présentés dans cette étude pourront être validés avec la technique du « gel mobility shift assay ». Cette méthode permet de vérifier si l'ARN est lié ou non à la protéine en évaluant le profil de migration de l'ARN sur gel. L'ARN lié aux protéines aura un profil de migration diffèrent de l'ARN non-lié. Ceci permettra de valider les interactions ARN-protéines identifiés par MS.

En conclusion, ces résultats démontrent qu'il est possible d'identifier des protéines qui lient au PAS de transcrits viraux à l'aide du système de purification par affinité à la streptavidine, et qu'avec quelques optimisations, cette technique pourrait être utilisée pour identifier les facteurs impliqués dans la polyadénylation alternative.

References

Anderson, K. P., Frink, R. J., Devi, G. B., Gaylord, B. H., Costa, R. H., and Wagner, E. K. (1981). Detailed characterization of the mRNA mapping in the HindIII fragment K region of the herpes simplex virus type 1 genome. *J Virol* **37**(3), 1011-27.

Aravind, L. (1999). An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico Biol* **1**(2), 69-91.

Arii, J., Goto, H., Suenaga, T., Oyama, M., Kozuka-Hata, H., Imai, T., Minowa, A., Akashi, H., Arase, H., Kawaoka, Y., and Kawaguchi, Y. (2010). Non-muscle myosin IIA is a functional entry receptor for herpes simplex virus-1. *Nature* **467**(7317), 859-62.

Atanasiu, D., Saw, W. T., Cohen, G. H., and Eisenberg, R. J. (2010). Cascade of events governing cell-cell fusion induced by herpes simplex virus glycoproteins gD, gH/gL, and gB. *J Virol* **84**(23), 12292-9.

Balbo, P. B., and Bohm, A. (2007). Mechanism of poly(A) polymerase: structure of the enzyme-MgATP-RNA ternary complex and kinetic analysis. *Structure* **15**(9), 1117-31.

Balbo, P. B., Meinke, G., and Bohm, A. (2005). Kinetic studies of yeast polyA polymerase indicate an induced fit mechanism for nucleotide specificity. *Biochemistry* **44**(21), 7777-86.

Balbo, P. B., Toth, J., and Bohm, A. (2007). X-ray crystallographic and steady state fluorescence characterization of the protein dynamics of yeast polyadenylate polymerase. *J Mol Biol* **366**(5), 1401-15.

Barabino, S. M., Hubner, W., Jenny, A., Minvielle-Sebastia, L., and Keller, W. (1997). The 30-kD subunit of mammalian cleavage and polyadenylation specificity factor and its yeast homolog are RNA-binding zinc finger proteins. *Genes Dev* **11**(13), 1703-16.

Barabino, S. M., Ohnacker, M., and Keller, W. (2000). Distinct roles of two Yth1p domains in 3'-end cleavage and polyadenylation of yeast pre-mRNAs. *EMBO J* **19**(14), 3778-87.

Bardwell, V. J., and Wickens, M. (1990). Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res* **18**(22), 6587-94.

Barilla, D., Lee, B. A., and Proudfoot, N. J. (2001). Cleavage/polyadenylation factor IA associates with the carboxyl-terminal domain of RNA polymerase II in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **98**(2), 445-50.

Beaudoing, E., Freier, S., Wyatt, J. R., Claverie, J. M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. *Genome Res* **10**(7), 1001-10.

Bernstein, D. S., Buter, N., Stumpf, C., and Wickens, M. (2002). Analyzing mRNA-protein complexes using a yeast three-hybrid system. *Methods* **26**(2), 123-41.

Bertrand, L., Leiva-Torres, G. A., Hyjazie, H., and Pearson, A. (2010). Conserved residues in the UL24 protein of herpes simplex virus 1 are important for dispersal of the nucleolar protein nucleolin. *J Virol* **84**(1), 109-18.

Bertrand, L., and Pearson, A. (2008). The conserved N-terminal domain of herpes simplex virus 1 UL24 protein is sufficient to induce the spatial redistribution of nucleolin. *J Gen Virol* **89**(Pt 5), 1142-51.

Blencowe, B. J., Sproat, B. S., Ryder, U., Barabino, S., and Lamond, A. I. (1989). Antisense probing of the human U4/U6 snRNP with biotinylated 2'-OMe RNA oligonucleotides. *Cell* **59**(3), 531-9.

Booy, F. P., Trus, B. L., Davison, A. J., and Steven, A. C. (1996). The capsid architecture of channel catfish virus, an evolutionarily distant herpesvirus, is largely conserved in the absence of discernible sequence homology with herpes simplex virus. *Virology* **215**(2), 134-41.

Bryant, H. E., Wadd, S. E., Lamond, A. I., Silverstein, S. J., and Clements, J. B. (2001). Herpes simplex virus IE63 (ICP27) protein interacts with spliceosome-associated protein 145 and inhibits splicing prior to the first catalytic step. *J Virol* **75**(9), 4376-85.

Callebaut, I., Moshous, D., Mornon, J. P., and de Villartay, J. P. (2002). Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res* **30**(16), 3592-601.

Cammas, A., Lewis, S. M., Vagner, S., and Holcik, M. (2008). Post-transcriptional control of gene expression through subcellular relocalization of mRNA binding proteins. *Biochem Pharmacol* **76**(11), 1395-403.

Campbell, M. E., Palfreyman, J. W., and Preston, C. M. (1984). Identification of herpes simplex virus DNA sequences which encode a trans-acting polypeptide responsible for stimulation of immediate early transcription. *J Mol Biol* **180**(1), 1-19.

Carr-Schmid, A., Durko, N., Cavallius, J., Merrick, W. C., and Kinzy, T. G. (1999). Mutations in a GTP-binding motif of eukaryotic elongation factor 1A reduce both translational fidelity and the requirement for nucleotide exchange. *J Biol Chem* **274**(42), 30297-302.

Chao, L. C., Jamil, A., Kim, S. J., Huang, L., and Martinson, H. G. (1999). Assembly of the cleavage and polyadenylation apparatus requires about 10 seconds in vivo and is faster for strong than for weak poly(A) sites. *Mol Cell Biol* **19**(8), 5588-600.

Charbonneau, M. E., Girard, V., Nikolakakis, A., Campos, M., Berthiaume, F., Dumas, F., Lepine, F., and Mourez, M. (2007). O-linked glycosylation ensures the normal conformation of the autotransporter adhesin involved in diffuse adherence. *J Bacteriol* **189**(24), 8880-9.

Colgan, D. F., Murthy, K. G., Prives, C., and Manley, J. L. (1996). Cell-cycle related regulation of poly(A) polymerase by phosphorylation. *Nature* **384**(6606), 282-5.

Cook, W. J., and Coen, D. M. (1996). Temporal regulation of herpes simplex virus type 1 UL24 mRNA expression via differential polyadenylation. *Virology* **218**(1), 204-13.

Cook, W. J., Lin, S. M., DeLuca, N. A., and Coen, D. M. (1995). Initiator elements and regulated expression of the herpes simplex virus thymidine kinase gene. *J Virol* **69**(11), 7291-4.

Cook, W. J., Wobbe, K. K., Boni, J., and Coen, D. M. (1996). Regulation of neighboring gene expression by the herpes simplex virus type 1 thymidine kinase gene. *Virology* **218**(1), 193-203.

Costessi, L., Devescovi, G., Baralle, F. E., and Muro, A. F. (2006). Brain-specific promoter and polyadenylation sites of the beta-adducin pre-mRNA generate an unusually long 3'-UTR. *Nucleic Acids Res* **34**(1), 243-53.

Crute, J. J., Tsurumi, T., Zhu, L. A., Weller, S. K., Olivo, P. D., Challberg, M. D., Mocarski, E. S., and Lehman, I. R. (1989). Herpes simplex virus 1 helicase-primase: a complex of three herpes-encoded gene products. *Proc Natl Acad Sci U S A* **86**(7), 2186-9.

Danckwardt, S., Hentze, M. W., and Kulozik, A. E. (2008). 3' end mRNA processing: molecular mechanisms and implications for health and disease. *EMBO J* **27**(3), 482-98.

Dantonel, J. C., Murthy, K. G., Manley, J. L., and Tora, L. (1997). Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature* **389**(6649), 399-402.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. *Mol Cell* **5**(5), 779-87.

Das, S., Vasanji, A., and Pellett, P. E. (2007). Three-dimensional structure of the human cytomegalovirus cytoplasmic virion assembly complex includes a reoriented secretory apparatus. *J Virol* **81**(21), 11861-9.

Dass, B., McMahon, K. W., Jenkins, N. A., Gilbert, D. J., Copeland, N. G., and MacDonald, C. C. (2001). The gene for a variant form of the polyadenylation protein CstF-64 is on chromosome 19 and is expressed in pachytene spermatocytes in mice. *J Biol Chem* **276**(11), 8044-50.

Dass, B., Tardif, S., Park, J. Y., Tian, B., Weitlauf, H. M., Hess, R. A., Carnes, K., Griswold, M. D., Small, C. L., and Macdonald, C. C. (2007). Loss of polyadenylation protein tauCstF-64 causes spermatogenic defects and male infertility. *Proc Natl Acad Sci U S A* **104**(51), 20374-9.

Davar, G., Kramer, M. F., Garber, D., Roca, A. L., Andersen, J. K., Bebrin, W., Coen, D. M., Kosz-Vnenchak, M., Knipe, D. M., Breakefield, X. O., and et al. (1994). Comparative efficacy of expression of genes delivered to mouse sensory neurons with herpes virus vectors. *J Comp Neurol* **339**(1), 3-11.

Davison, A. J. (2002). Evolution of the herpesviruses. *Vet Microbiol* **86**(1-2), 69-88.

Davison, A. J., Eberle, R., Ehlers, B., Hayward, G. S., McGeoch, D. J., Minson, A. C., Pellett, P. E., Roizman, B., Studdert, M. J., and Thiry, E. (2009). The order Herpesvirales. *Arch Virol* **154**(1), 171-7.

de Bruyn Kops, A., and Knipe, D. M. (1988). Formation of DNA replication structures in herpes virus-infected cells requires a viral DNA binding protein. *Cell* **55**(5), 857-68.

de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H., and Keller, W. (2000). Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J* **19**(21), 5895-904.

Deiss, L. P., Chou, J., and Frenkel, N. (1986). Functional domains within the a sequence involved in the cleavage-packaging of herpes simplex virus DNA. *J Virol* **59**(3), 605-18.

Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999). Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**(6), 835-45.

Desai, P., and Person, S. (1996). Molecular interactions between the HSV-1 capsid proteins as measured by the yeast two-hybrid system. *Virology* **220**(2), 516-21.

Desselberger, U. (1998). Herpes simplex virus infection in pregnancy: diagnosis and significance. *Intervirology* **41**(4-5), 185-90.

Dettwiler, S., Aringhieri, C., Cardinale, S., Keller, W., and Barabino, S. M. (2004). Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding, protein-protein interactions, and subcellular localization. *J Biol Chem* **279**(34), 35788-97.

Dobrikova, E., Shveygert, M., Walters, R., and Gromeier, M. (2010). Herpes simplex virus proteins ICP27 and UL47 associate with polyadenylate-binding protein and control its subcellular distribution. *J Virol* **84**(1), 270-9.

Dominski, Z. (2007). Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. *Crit Rev Biochem Mol Biol* **42**(2), 67-93.

Dreyfuss, G., Kim, V. N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol* **3**(3), 195-205.

Edmonds, M. (2002). A history of poly A sequences: from formation to factors to function. *Prog Nucleic Acid Res Mol Biol* **71**, 285-389.

Edmonds, M., and Abrams, R. (1960). Polynucleotide biosynthesis: formation of a sequence of adenylate units from adenosine triphosphate by an enzyme from thymus nuclei. *J Biol Chem* **235**, 1142-9.

Epstein, A. L. (2009). Progress and prospects: biological properties and technological advances of herpes simplex virus type 1-based amplicon vectors. *Gene Ther* **16**(6), 709-15.

Epstein, M. A. (1962). Observations on the mode of release of herpes virus from infected HeLa cells. *J Cell Biol* **12,** 589-97.

Ford, L. P., Bagga, P. S., and Wilusz, J. (1997). The poly(A) tail inhibits the assembly of a 3'-to-5' exonuclease in an in vitro RNA stability system. *Mol Cell Biol* **17**(1), 398-406.

Fribourg, S., Gatfield, D., Izaurralde, E., and Conti, E. (2003). A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat Struct Biol* **10**(6), 433-9.

Fruh, K., Ahn, K., Djaballah, H., Sempe, P., van Endert, P. M., Tampe, R., Peterson, P. A., and Yang, Y. (1995). A viral inhibitor of peptide transporters for antigen presentation. *Nature* **375**(6530), 415-8.

Furlong, D., Swift, H., and Roizman, B. (1972). Arrangement of herpesvirus deoxyribonucleic acid in the core. *J Virol* **10**(5), 1071-4.

Gaddy, C. E., Wong, D. S., Markowitz-Shulman, A., and Colberg-Poley, A. M. (2010). Regulation of the subcellular distribution of key cellular RNA-processing factors during permissive human cytomegalovirus infection. *J Gen Virol* **91**(Pt 6), 1547-59.

Garber, D. A., Beverley, S. M., and Coen, D. M. (1993). Demonstration of circularization of herpes simplex virus DNA following infection using pulsed field gel electrophoresis. *Virology* **197**(1), 459-62.

Goins, W. F., and Stinski, M. F. (1986). Expression of a human cytomegalovirus late gene is posttranscriptionally regulated by a 3'-end-processing event occurring exclusively late after infection. *Mol Cell Biol* **6**(12), 4202-13.

Gottlieb, J., Marcy, A. I., Coen, D. M., and Challberg, M. D. (1990). The herpes simplex virus type 1 UL42 gene product: a subunit of DNA polymerase that functions to increase processivity. *J Virol* **64**(12), 5976-87.

Grosshans, H., Hurt, E., and Simos, G. (2000). An aminoacylation-dependent nuclear tRNA export pathway in yeast. *Genes Dev* **14**(7), 830-40.

Gu, B., and DeLuca, N. (1994). Requirements for activation of the herpes simplex virus glycoprotein C promoter in vitro by the viral regulatory protein ICP4. *J Virol* **68**(12), 7953-65.

Guzowski, J. F., Singh, J., and Wagner, E. K. (1994). Transcriptional activation of the herpes simplex virus type 1 UL38 promoter conferred by the cis-acting downstream activation sequence is mediated by a cellular transcription factor. *J Virol* **68**(12), 7774-89.

Guzowski, J. F., and Wagner, E. K. (1993). Mutational analysis of the herpes simplex virus type 1 strict late UL38 promoter/leader reveals two regions critical in transcriptional regulation. *J Virol* **67**(9), 5098-108.

Hann, L. E., Cook, W. J., Uprichard, S. L., Knipe, D. M., and Coen, D. M. (1998). The role of herpes simplex virus ICP27 in the regulation of UL24 gene expression by differential polyadenylation. *J Virol* **72**(10), 7709-14.

Harrison, S. C. (2008). Viral membrane fusion. *Nat Struct Mol Biol* **15**(7), 690-8.

He, X., and Moore, C. (2005). Regulation of yeast mRNA 3' end processing by phosphorylation. *Mol Cell* **19**(5), 619-29.

Heldwein, E. E., and Krummenacher, C. (2008). Entry of herpesviruses into mammalian cells. *Cell Mol Life Sci* **65**(11), 1653-68.

Herold, B. C., Visalli, R. J., Susmarski, N., Brandt, C. R., and Spear, P. G. (1994). Glycoprotein C-independent binding of herpes simplex virus to cells requires cell surface heparan sulphate and glycoprotein B. *J Gen Virol* **75 (Pt 6),** 1211-22.

Herold, B. C., WuDunn, D., Soltys, N., and Spear, P. G. (1991). Glycoprotein C of herpes simplex virus type 1 plays a principal role in the adsorption of virus to cells and in infectivity. *J Virol* **65**(3), 1090-8.

Hirose, Y., and Manley, J. L. (1998). RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395**(6697), 93-6.

Hogg, J. R., and Collins, K. (2007). RNA-based affinity purification reveals 7SK RNPs with distinct composition and regulation. *RNA* **13**(6), 868-80.

Hollingworth, D., Noble, C. G., Taylor, I. A., and Ramos, A. (2006). RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11. *RNA* **12**(4), 555-60.

Holtke, H. J., Ankenbauer, W., Muhlegger, K., Rein, R., Sagner, G., Seibl, R., and Walter, T. (1995). The digoxigenin (DIG) system for non-radioactive labelling and detection of nucleic acids--an overview. *Cell Mol Biol (Noisy-le-grand)* **41**(7), 883-905.

Honess, R. W., and Roizman, B. (1974). Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *J Virol* **14**(1), 8-19.

Hu, J., Lutz, C. S., Wilusz, J., and Tian, B. (2005). Bioinformatic identification of candidate cis-regulatory elements involved in human mRNA polyadenylation. *RNA* **11**(10), 1485-93.

Hughes, T. A., La Boissiere, S., and O'Hare, P. (1999). Analysis of functional domains of the host cell factor involved in VP16 complex formation. *J Biol Chem* **274**(23), 16437-43.

Hyde-DeRuyscher, R. P., and Carmichael, G. G. (1990). Polyomavirus late pre-mRNA processing: DNA replication-associated changes in leader exon multiplicity suggest a role for leader-to-leader splicing in the early-late switch. *J Virol* **64**(12), 5823-32.

Ingram, A., Phelan, A., Dunlop, J., and Clements, J. B. (1996). Immediate early protein IE63 of herpes simplex virus type 1 binds RNA directly. *J Gen Virol* **77 (Pt 8)**, 1847-51.

Jacob, R. J., Morse, L. S., and Roizman, B. (1979). Anatomy of herpes simplex virus DNA. XII. Accumulation of head-to-tail concatemers in nuclei of infected cells and their role in the generation of the four isomeric arrangements of viral DNA. *J Virol* **29**(2), 448-57.

Jacobs, A., Breakefield, X. O., and Fraefel, C. (1999). HSV-1-based vectors for gene therapy of neurological diseases and brain tumors: part I. HSV-1 structure, replication and pathogenesis. *Neoplasia* **1**(5), 387-401.

Jacobson, J. G., Chen, S. H., Cook, W. J., Kramer, M. F., and Coen, D. M. (1998). Importance of the herpes simplex virus UL24 gene for productive ganglionic infection in mice. *Virology* **242**(1), 161-9.

Jacobson, J. G., Martin, S. L., and Coen, D. M. (1989). A conserved open reading frame that overlaps the herpes simplex virus thymidine kinase gene is important for viral growth in cell culture. *J Virol* **63**(4), 1839-43.

Jenny, A., Hauri, H. P., and Keller, W. (1994). Characterization of cleavage and polyadenylation specificity factor and cloning of its 100-kilodalton subunit. *Mol Cell Biol* **14**(12), 8183-90.

Johnson, L. A., and Sandri-Goldin, R. M. (2009). Efficient nuclear export of herpes simplex virus 1 transcripts requires both RNA binding by ICP27 and ICP27 interaction with TAP/NXF1. *J Virol* **83**(3), 1184-92.

Jones, P. C., and Roizman, B. (1979). Regulation of herpesvirus macromolecular synthesis. VIII. The transcription program consists of three phases during which both extent of transcription and accumulation of RNA in the cytoplasm are regulated. *J Virol* **31**(2), 299-314.

Kahvejian, A., Roy, G., and Sonenberg, N. (2001). The mRNA closed-loop model: the function of PABP and PABP-interacting proteins in mRNA translation. *Cold Spring Harb Symp Quant Biol* **66**, 293-300.

Keller, R. W., Kuhn, U., Aragon, M., Bornikova, L., Wahle, E., and Bear, D. G. (2000). The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. *J Mol Biol* **297**(3), 569-83.

Kennard, J., Rixon, F. J., McDougall, I. M., Tatman, J. D., and Preston, V. G. (1995). The 25 amino acid residues at the carboxy terminus of the herpes simplex virus type 1 UL26.5 protein are required for the formation of the capsid shell around the scaffold. *J Gen Virol* **76 (Pt 7)**, 1611-21.

Kennedy, I. M., Haddow, J. K., and Clements, J. B. (1990). Analysis of human papillomavirus type 16 late mRNA 3' processing signals in vitro and in vivo. *J Virol* **64**(4), 1825-9.

Kent, J. R., Zeng, P. Y., Atanasiu, D., Gardner, J., Fraser, N. W., and Berger, S. L. (2004). During lytic infection herpes simplex virus type 1 is associated with histones bearing modifications that correlate with active transcription. *J Virol* **78**(18), 10178-86.

Kerwitz, Y., Kuhn, U., Lilie, H., Knoth, A., Scheuermann, T., Friedrich, H., Schwarz, E., and Wahle, E. (2003). Stimulation of poly(A) polymerase through a direct interaction with the nuclear poly(A) binding protein allosterically regulated by RNA. *EMBO J* **22**(14), 3705-14.

Khacho, M., Mekhail, K., Pilon-Larose, K., Pause, A., Cote, J., and Lee, S. (2008). eEF1A is a novel component of the mammalian nuclear protein export machinery. *Mol Biol Cell* **19**(12), 5296-308.

Kibler, P. K., Duncan, J., Keith, B. D., Hupel, T., and Smiley, J. R. (1991). Regulation of herpes simplex virus true late gene expression: sequences downstream from the US11 TATA box inhibit expression from an unreplicated template. *J Virol* **65**(12), 6749-60.

Kieff, E., Hoyer, B., Bachenheimer, S., and Roizman, B. (1972). Genetic relatedness of type 1 and type 2 herpes simplex viruses. *J Virol* **9**(5), 738-45.

Kim, D. B., Zabierowski, S., and DeLuca, N. A. (2002). The initiator element in a herpes simplex virus type 1 late-gene promoter enhances activation by ICP4, resulting in abundant late-gene expression. *J Virol* **76**(4), 1548-58.

Kim, H., and Lee, Y. (2001). Interaction of poly(A) polymerase with the 25-kDa subunit of cleavage factor I. *Biochem Biophys Res Commun* **289**(2), 513-8.

Kimberlin, D. W. (2004). Neonatal herpes simplex infection. *Clin Microbiol Rev* **17**(1), 1-13.

Knipe, D. M., and Cliffe, A. (2008). Chromatin control of herpes simplex virus lytic and latent infection. *Nat Rev Microbiol* **6**(3), 211-21.

Knipe, D. M., Senechek, D., Rice, S. A., and Smith, J. L. (1987). Stages in the nuclear association of the herpes simplex virus transcriptional activator protein ICP4. *J Virol* **61**(2), 276-84.

Kuersten, S., and Goodwin, E. B. (2003). The power of the 3' UTR: translational control and development. *Nat Rev Genet* 4(8), 626-37.

Kuhn, U., and Wahle, E. (2004). Structure and function of poly(A) binding proteins. *Biochim Biophys Acta* **1678**(2-3), 67-84.

Kwong, A. D., Kruper, J. A., and Frenkel, N. (1988). Herpes simplex virus virion host shutoff function. *J Virol* **62**(3), 912-21.

Kyburz, A., Sadowski, M., Dichtl, B., and Keller, W. (2003). The role of the yeast cleavage and polyadenylation factor subunit Ydh1p/Cft2p in pre-mRNA 3'-end formation. *Nucleic Acids Res* **31**(14), 3936-45.

LaBoissiere, S., and O'Hare, P. (2000). Analysis of HCF, the cellular cofactor of VP16, in herpes simplex virus-infected cells. *J Virol* **74**(1), 99-109.

Le Hir, H., and Seraphin, B. (2008). EJCs at the heart of translational control. *Cell* **133**(2), 213-6.

Leiva-Torres, G. A., Rochette, P. A., and Pearson, A. (2010). Differential importance of highly conserved residues in UL24 for herpes simplex virus 1 replication in vivo and reactivation. *J Gen Virol* **91**(Pt 5), 1109-16.

Leonov, A. A., Sergiev, P. V., Bogdanov, A. A., Brimacombe, R., and Dontsova, O. A. (2003). Affinity purification of ribosomes with a lethal G2655C mutation in 23 S rRNA that affects the translocation. *J Biol Chem* **278**(28), 25664-70.

Liang, S., and Lutz, C. S. (2006). p54nrb is a component of the snRNP-free U1A (SF-A) complex that promotes pre-mRNA cleavage during polyadenylation. *RNA* **12**(1), 111-21.

Liao, H. J., Kobayashi, R., and Mathews, M. B. (1998). Activities of adenovirus virus-associated RNAs: purification and characterization of RNA binding proteins. *Proc Natl Acad Sci U S A* **95**(15), 8514-9.

Lingner, J., and Cech, T. R. (1996). Purification of telomerase from Euplotes aediculatus: requirement of a primer 3' overhang. *Proc Natl Acad Sci U S A* **93**(20), 10712-7.

Liu, F. Y., and Roizman, B. (1991). The herpes simplex virus 1 gene encoding a protease also contains within its coding domain the gene encoding the more abundant substrate. *J Virol* **65**(10), 5149-56.

Lutz, C. S. (2008). Alternative polyadenylation: a twist on mRNA 3' end formation. *ACS Chem Biol* **3**(10), 609-17.

Lymberopoulos, M. H., and Pearson, A. (2007). Involvement of UL24 in herpes-simplex-virus-1-induced dispersal of nucleolin. *Virology* **363**(2), 397-409.

Mandel, C. R., Bai, Y., and Tong, L. (2008). Protein factors in pre-mRNA 3'-end processing. *Cell Mol Life Sci* **65**(7-8), 1099-122.

Mandel, C. R., Kaneko, S., Zhang, H., Gebauer, D., Vethantham, V., Manley, J. L., and Tong, L. (2006). Polyadenylation factor CPSF-73 is the pre-mRNA 3'-end-processing endonuclease. *Nature* **444**(7121), 953-6.

Mandel, C. R., and Tong, L. (2007). How to get all "A"s in polyadenylation. *Structure* **15**(9), 1024-6.

Mann, K. P., Weiss, E. A., and Nevins, J. R. (1993). Alternative poly(A) site utilization during adenovirus infection coincides with a decrease in the activity of a poly(A) site processing factor. *Mol Cell Biol* **13**(4), 2411-9.

Martin, T. E., Barghusen, S. C., Leser, G. P., and Spear, P. G. (1987). Redistribution of nuclear ribonucleoprotein antigens during herpes simplex virus infection. *J Cell Biol* **105**(5), 2069-82.

Martincic, K., Alkan, S. A., Cheatle, A., Borghesi, L., and Milcarek, C. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nat Immunol* **10**(10), 1102-9.

Martincic, K., Campbell, R., Edwalds-Gilbert, G., Souan, L., Lotze, M. T., and Milcarek, C. (1998). Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G0 to S phase transition. *Proc Natl Acad Sci U S A* **95**(19), 11095-100.

Maul, G. G., Guldner, H. H., and Spivack, J. G. (1993). Modification of discrete nuclear domains induced by herpes simplex virus type 1 immediate early gene 1 product (ICP0). *J Gen Virol* **74 (Pt 12)**, 2679-90.

Mauxion, F., Chen, C. Y., Seraphin, B., and Shyu, A. B. (2009). BTG/TOB factors impact deadenylases. *Trends Biochem Sci* **34**(12), 640-7.

McClain, D. S., and Fuller, A. O. (1994). Cell-specific kinetics and efficiency of herpes simplex virus type 1 entry are determined by two distinct phases of attachment. *Virology* **198**(2), 690-702.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S. D., Wickens, M., and Bentley, D. L. (1997). The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385**(6614), 357-61.

McGeoch, D. J., Dalrymple, M. A., Davison, A. J., Dolan, A., Frame, M. C., McNab, D., Perry, L. J., Scott, J. E., and Taylor, P. (1988). The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J Gen Virol* **69 (Pt 7),** 1531-74.

McGregor, F., Phelan, A., Dunlop, J., and Clements, J. B. (1996). Regulation of herpes simplex virus poly (A) site usage and the action of immediate-early protein IE63 in the early-late switch. *J Virol* **70**(3), 1931-40.

McNab, A. R., Desai, P., Person, S., Roof, L. L., Thomsen, D. R., Newcomb, W. W., Brown, J. C., and Homa, F. L. (1998). The product of the herpes simplex virus type 1 UL25 gene is required for encapsidation but not for cleavage of replicated viral DNA. *J Virol* **72**(2), 1060-70.

Meagher, M. J., Schumacher, J. M., Lee, K., Holdcraft, R. W., Edelhoff, S., Disteche, C., and Braun, R. E. (1999). Identification of ZFR, an ancient and highly conserved murine chromosome-associated zinc finger protein. *Gene* **228**(1-2), 197-211.

Mettenleiter, T. C. (2006). Intriguing interplay between viral proteins during herpesvirus assembly or: the herpesvirus assembly puzzle. *Vet Microbiol* **113**(3-4), 163-9.

Mettenleiter, T. C., Klupp, B. G., and Granzow, H. (2006). Herpesvirus assembly: a tale of two membranes. *Curr Opin Microbiol* **9**(4), 423-9.

Mettenleiter, T. C., Klupp, B. G., and Granzow, H. (2009). Herpesvirus assembly: an update. *Virus Res* **143**(2), 222-34.

Meyer, S., Urbanke, C., and Wahle, E. (2002). Equilibrium studies on the association of the nuclear poly(A) binding protein with poly(A) of different lengths. *Biochemistry* **41**(19), 6082-9.

Millevoi, S., and Vagner, S. (2010). Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* **38**(9), 2757-74.

Monier, K., Armas, J. C., Etteldorf, S., Ghazal, P., and Sullivan, K. F. (2000). Annexation of the interchromosomal space during viral infection. *Nat Cell Biol* **2**(9), 661-5.

Negrutskii, B. S., and El'skaya, A. V. (1998). Eukaryotic translation elongation factor 1 alpha: structure, expression, functions, and possible role in aminoacyl-tRNA channeling. *Prog Nucleic Acid Res Mol Biol* **60,** 47-78.

Nevins, J. R., and Wilson, M. C. (1981). Regulation of adenovirus-2 gene expression at the level of transcriptional termination and RNA processing. *Nature* **290**(5802), 113-8.

Newcomb, W. W., and Brown, J. C. (1994). Induced extrusion of DNA from the capsid of herpes simplex virus type 1. *J Virol* **68**(1), 433-40.

Newcomb, W. W., Homa, F. L., and Brown, J. C. (2006). Herpes simplex virus capsid structure: DNA packaging protein UL25 is located on the external surface of the capsid near the vertices. *J Virol* **80**(13), 6286-94.

Newcomb, W. W., Trus, B. L., Booy, F. P., Steven, A. C., Wall, J. S., and Brown, J. C. (1993). Structure of the herpes simplex virus capsid. Molecular composition of the pentons and the triplexes. *J Mol Biol* **232**(2), 499-511.

Olivo, P. D., Nelson, N. J., and Challberg, M. D. (1988). Herpes simplex virus DNA replication: the UL9 gene encodes an origin-binding protein. *Proc Natl Acad Sci U S A* **85**(15), 5414-8.

Parrott, A. M., Walsh, M. R., Reichman, T. W., and Mathews, M. B. (2005). RNA binding and phosphorylation determine the intracellular distribution of nuclear factors 90 and 110. *J Mol Biol* **348**(2), 281-93.

Pearson, A., and Coen, D. M. (2002). Identification, localization, and regulation of expression of the UL24 protein of herpes simplex virus type 1. *J Virol* **76**(21), 10821-8.

Pearson, A., Knipe, D. M., and Coen, D. M. (2004). ICP27 selectively regulates the cytoplasmic localization of a subset of viral transcripts in herpes simplex virus type 1-infected cells. *J Virol* **78**(1), 23-32.

Pellett, P. E., and Roizman, B. (2007). The Family *Herpesviridae*: A Brief Introduction. 5th ed. *In* "Fields Virology" (D. M. Knipe, Ed.), Vol. 2, pp. 2479-2499. Lippincott Williams & Wilkins, Philadelphia.

Phelan, A., Dunlop, J., and Clements, J. B. (1996). Herpes simplex virus type 1 protein IE63 affects the nuclear export of virus intron-containing transcripts. *J Virol* **70**(8), 5255-65.

Porath, J. (1992). Immobilized metal ion affinity chromatography. *Protein Expr Purif* **3**(4), 263-81.

Porath, J., Carlsson, J., Olsson, I., and Belfrage, G. (1975). Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* **258**(5536), 598-9.

Post, L. E., Mackem, S., and Roizman, B. (1981). Regulation of alpha genes of herpes simplex virus: expression of chimeric genes produced by fusion of thymidine kinase with alpha gene promoters. *Cell* **24**(2), 555-65.

Preston, V. G., Rixon, F. J., McDougall, I. M., McGregor, M., and al Kobaisi, M. F. (1992). Processing of the herpes simplex virus assembly protein ICP35 near its carboxy terminal end requires the product of the whole of the UL26 reading frame. *Virology* **186**(1), 87-98.

Prickett, K. S., Amberg, D. C., and Hopp, T. P. (1989). A calcium-dependent antibody for identification and purification of recombinant proteins. *Biotechniques* **7**(6), 580-9.

Proudfoot, N. (2004). New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* **16**(3), 272-8.

Proudnikov, D., and Mirzabekov, A. (1996). Chemical methods of DNA and RNA fluorescent labeling. *Nucleic Acids Res* **24**(22), 4535-42.

Quinlan, M. P., Chen, L. B., and Knipe, D. M. (1984). The intranuclear location of a herpes simplex virus DNA-binding protein is determined by the status of viral DNA replication. *Cell* **36**(4), 857-68.

Rajcani, J., Andrea, V., and Ingeborg, R. (2004). Peculiarities of herpes simplex virus (HSV) transcription: an overview. *Virus Genes* **28**(3), 293-310.

Read, G. S., Sharp, J. A., and Summers, W. C. (1984). In vitro and in vivo transcription initiation sites on the TK-encoding BamHI Q fragment of HSV-1 DNA. *Virology* **138**(2), 368-72.

Read, G. S., and Summers, W. C. (1982). In vitro transcription of the thymidine kinase gene of herpes simplex virus. *Proc Natl Acad Sci U S A* **79**(17), 5215-9.

Reed, V. S., Wastney, M. E., and Yang, D. C. (1994). Mechanisms of the transfer of aminoacyl-tRNA from aminoacyl-tRNA synthetase to the elongation factor 1 alpha. *J Biol Chem* **269**(52), 32932-6.

Reichman, T. W., and Mathews, M. B. (2003). RNA binding and intramolecular interactions modulate the regulation of gene expression by nuclear factor 110. *RNA* **9**(5), 543-54.

Reynolds, A. E., Liang, L., and Baines, J. D. (2004). Conformational changes in the nuclear lamina induced by herpes simplex virus type 1 require genes U(L)31 and U(L)34. *J Virol* **78**(11), 5564-75.

Roget, A., Bazin, H., and Teoule, R. (1989). Synthesis and use of labelled nucleoside phosphoramidite building blocks bearing a reporter group: biotinyl, dinitrophenyl, pyrenyl and dansyl. *Nucleic Acids Res* **17**(19), 7643-51.

Roizman, B. (1979). The structure and isomerization of herpes simplex virus genomes. *Cell* **16**(3), 481-94.

Roizman, B. (1996). The function of herpes simplex virus genes: a primer for genetic engineering of novel vectors. *Proc Natl Acad Sci U S A* **93**(21), 11307-12.

Roizman, B., Knipe, D. M., and Whitley, R. J. (2007). Herpes Simplex Viruses. 5th ed. *In* "Fields Virology" (D. M. Knipe, and P. M. Howley, Eds.), Vol. 2. Lippincott Williams & Wilkins, Philadelphia.

Rouault, T. A., Hentze, M. W., Haile, D. J., Harford, J. B., and Klausner, R. D. (1989). The iron-responsive element binding protein: a method for the affinity purification of a regulatory RNA-binding protein. *Proc Natl Acad Sci U S A* **86**(15), 5768-72.

Ruegsegger, U., Beyer, K., and Keller, W. (1996). Purification and characterization of human cleavage factor Im involved in the 3' end processing of messenger RNA precursors. *J Biol Chem* **271**(11), 6107-13.

Ruegsegger, U., Blank, D., and Keller, W. (1998). Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol Cell* **1**(2), 243-53.

Ryan, K. (2007). Pre-mRNA 3' cleavage is reversibly inhibited in vitro by cleavage factor dephosphorylation. *RNA Biol* 4(1), 26-33.

Ryan, K., Calvo, O., and Manley, J. L. (2004). Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA* **10**(4), 565-73.

Sachs, A. B., Sarnow, P., and Hentze, M. W. (1997). Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell* **89**(6), 831-8.

Sanders, P. G., Wilkie, N. M., and Davison, A. J. (1982). Thymidine kinase deletion mutants of herpes simplex virus type 1. *J Gen Virol* **63**(2), 277-95.

Sandri-Goldin, R. M. (2003). Replication of the herpes simplex virus genome: does it really go around in circles? *Proc Natl Acad Sci U S A* **100**(13), 7428-9.

Schleiss, M. R. (2009). Persistent and recurring viral infections: the human herpesviruses. *Curr Probl Pediatr Adolesc Health Care* **39**(1), 7-23.

Sheldrick, P., and Berthelot, N. (1975). Inverted repetitions in the chromosome of herpes simplex virus. *Cold Spring Harb Symp Quant Biol* **39 Pt 2**, 667-78.

Shi, L., Zhao, G., Qiu, D., Godfrey, W. R., Vogel, H., Rando, T. A., Hu, H., and Kao, P. N. (2005). NF90 regulates cell cycle exit and terminal myogenic differentiation by direct binding to the 3'-untranslated region of MyoD and p21WAF1/CIP1 mRNAs. *J Biol Chem* **280**(19), 18981-9.

Shilatifard, A., Duan, D. R., Haque, D., Florence, C., Schubach, W. H., Conaway, J. W., and Conaway, R. C. (1997a). ELL2, a new member of an ELL family of RNA polymerase II elongation factors. *Proc Natl Acad Sci U S A* **94**(8), 3639-43.

Shilatifard, A., Haque, D., Conaway, R. C., and Conaway, J. W. (1997b). Structure and function of RNA polymerase II elongation factor ELL. Identification of two overlapping ELL functional domains that govern its interaction with polymerase and the ternary elongation complex. *J Biol Chem* **272**(35), 22355-63.

Shim, J., Lim, H., J, R. Y., and Karin, M. (2002). Nuclear export of NF90 is required for interleukin-2 mRNA stabilization. *Mol Cell* **10**(6), 1331-44.

Shimazu, T., Horinouchi, S., and Yoshida, M. (2007). Multiple histone deacetylases and the CREB-binding protein regulate pre-mRNA 3'-end processing. *J Biol Chem* **282**(7), 4470-8.

Simons, P. C., and Vander Jagt, D. L. (1981). Purification of glutathione S-transferases by glutathione-affinity chromatography. *Methods Enzymol* **77**, 235-7.

Singh, J., and Wagner, E. K. (1993). Transcriptional analysis of the herpes simplex virus type 1 region containing the TRL/UL junction. *Virology* **196**(1), 220-31.

Skaliter, R., Makhov, A. M., Griffith, J. D., and Lehman, I. R. (1996). Rolling circle DNA replication by extracts of herpes simplex virus type 1-infected human cells. *J Virol* **70**(2), 1132-6.

Sodeik, B., Ebersold, M. W., and Helenius, A. (1997). Microtubule-mediated transport of incoming herpes simplex virus 1 capsids to the nucleus. *J Cell Biol* **136**(5), 1007-21.

Sopta, M., Carthew, R. W., and Greenblatt, J. (1985). Isolation of three proteins that bind to mammalian RNA polymerase II. *J Biol Chem* **260**(18), 10353-60.

Spear, P. G., Eisenberg, R. J., and Cohen, G. H. (2000). Three classes of cell surface receptors for alphaherpesvirus entry. *Virology* **275**(1), 1-8.

Srisawat, C., and Engelke, D. R. (2001). Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* **7**(4), 632-41.

Srisawat, C., and Engelke, D. R. (2002). RNA affinity tags for purification of RNAs and ribonucleoprotein complexes. *Methods* **26**(2), 156-61.

Srisawat, C., Goldstein, I. J., and Engelke, D. R. (2001). Sephadex-binding RNA ligands: rapid affinity purification of RNA from complex RNA mixtures. *Nucleic Acids Res* **29**(2), E4.

Stamminger, T., Puchtler, E., and Fleckenstein, B. (1991). Discordant expression of the immediate-early 1 and 2 gene regions of human cytomegalovirus at early times after infection involves posttranscriptional processing events. *J Virol* **65**(5), 2273-82.

Stow, N. D., and McMonagle, E. C. (1983). Characterization of the TRS/IRS origin of DNA replication of herpes simplex virus type 1. *Virology* **130**(2), 427-38.

Strom, T., and Frenkel, N. (1987). Effects of herpes simplex virus on mRNA stability. *J Virol* **61**(7), 2198-207.

Taddeo, B., Esclatine, A., and Roizman, B. (2004). Post-transcriptional processing of cellular RNAs in herpes simplex virus-infected cells. *Biochem Soc Trans* **32**(Pt 5), 697-701.

Takagaki, Y., and Manley, J. L. (1994). A polyadenylation factor subunit is the human homologue of the Drosophila suppressor of forked protein. *Nature* **372**(6505), 471-4.

Takagaki, Y., and Manley, J. L. (1997). RNA recognition by the human polyadenylation factor CstF. *Mol Cell Biol* **17**(7), 3907-14.

Takagaki, Y., and Manley, J. L. (2000). Complex protein interactions within the human polyadenylation machinery identify a novel component. *Mol Cell Biol* **20**(5), 1515-25.

Takagaki, Y., Seipelt, R. L., Peterson, M. L., and Manley, J. L. (1996). The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**(5), 941-52.

Tatman, J. D., Preston, V. G., Nicholson, P., Elliott, R. M., and Rixon, F. J. (1994). Assembly of herpes simplex virus type 1 capsids using a panel of recombinant baculoviruses. *J Gen Virol* **75 (Pt 5)**, 1101-13.

Thomsen, D. R., Roof, L. L., and Homa, F. L. (1994). Assembly of herpes simplex virus (HSV) intermediate capsids in insect cells infected with recombinant baculoviruses expressing HSV capsid proteins. *J Virol* **68**(4), 2442-57.

Tian, B., Hu, J., Zhang, H., and Lutz, C. S. (2005). A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Res* **33**(1), 201-12.

Trus, B. L., Booy, F. P., Newcomb, W. W., Brown, J. C., Homa, F. L., Thomsen, D. R., and Steven, A. C. (1996). The herpes simplex virus procapsid: structure, conformational changes upon maturation, and roles of the triplex proteins VP19c and VP23 in assembly. *J Mol Biol* **263**(3), 447-62.

Turcotte, S., Letellier, J., and Lippé, R. (2005). Herpes simplex virus type 1 capsids transit by the trans-Golgi network, where viral glycoproteins accumulate independently of capsid egress. *J Virol* **79**(14), 8847-60.

Turner, A., Bruun, B., Minson, T., and Browne, H. (1998). Glycoproteins gB, gD, and gHgL of herpes simplex virus type 1 are necessary and sufficient to mediate membrane fusion in a Cos cell transfection system. *J Virol* **72**(1), 873-5.

Uprichard, S. L., and Knipe, D. M. (1996). Herpes simplex ICP27 mutant viruses exhibit reduced expression of specific DNA replication genes. *J Virol* **70**(3), 1969-80.

Vinciguerra, P., and Stutz, F. (2004). mRNA export: an assembly line from genes to nuclear pores. *Curr Opin Cell Biol* **16**(3), 285-92.

Wadsworth, S., Jacob, R. J., and Roizman, B. (1975). Anatomy of herpes simplex virus DNA. II. Size, composition, and arrangement of inverted terminal repetitions. *J Virol* **15**(6), 1487-97.

Wagner, E. K., and Bloom, D. C. (1997). Experimental investigation of herpes simplex virus latency. *Clin Microbiol Rev* **10**(3), 419-43.

Wagner, E. K., Guzowski, J. F., and Singh, J. (1995). Transcription of the herpes simplex virus genome during productive and latent infection. *Prog Nucleic Acid Res Mol Biol* **51**, 123-65.

Walker, S. C., Scott, F. H., Srisawat, C., and Engelke, D. R. (2008). RNA affinity tags for the rapid purification and investigation of RNAs and RNA-protein complexes. *Methods Mol Biol* **488**, 23-40.

Wallace, A. M., Denison, T. L., Attaya, E. N., and MacDonald, C. C. (2004). Developmental distribution of the polyadenylation protein CstF-64 and the variant tauCstF-64 in mouse and rat testis. *Biol Reprod* **70**(4), 1080-7.

Wang, Q. Y., Zhou, C., Johnson, K. E., Colgrove, R. C., Coen, D. M., and Knipe, D. M. (2005). Herpesviral latency-associated transcript gene promotes assembly of heterochromatin on viral lytic-gene promoters in latent infection. *Proc Natl Acad Sci U S A* **102**(44), 16055-9.

Weir, J. P. (2001). Regulation of herpes simplex virus gene expression. *Gene* **271**(2), 117-30.

Weller, S. K., Lee, K. J., Sabourin, D. J., and Schaffer, P. A. (1983). Genetic analysis of temperature-sensitive mutants which define the gene for the major herpes simplex virus type 1 DNA-binding protein. *J Virol* **45**(1), 354-66.

Weller, S. K., Spadaro, A., Schaffer, J. E., Murray, A. W., Maxam, A. M., and Schaffer, P. A. (1985). Cloning, sequencing, and functional analysis of oriL, a herpes simplex virus type 1 origin of DNA synthesis. *Mol Cell Biol* **5**(5), 930-42.

Wickens, M., Anderson, P., and Jackson, R. J. (1997). Life and death in the cytoplasm: messages from the 3' end. *Curr Opin Genet Dev* **7**(2), 220-32.

Wild, P., Engels, M., Senn, C., Tobler, K., Ziegler, U., Schraner, E. M., Loepfe, E., Ackermann, M., Mueller, M., and Walther, P. (2005). Impairment of nuclear pores in bovine herpesvirus 1-infected MDBK cells. *J Virol* **79**(2), 1071-83.

Wild, P., Schraner, E. M., Cantieni, D., Loepfe, E., Walther, P., Muller, M., and Engels, M. (2002). The significance of the Golgi complex in envelopment of bovine herpesvirus 1 (BHV-1) as revealed by cryobased electron microscopy. *Micron* **33**(4), 327-37.

Wildy, P., and Watson, D. H. (1962). Electron microscopic studies on the architecture of animal viruses. *Cold Spring Harb Symp Quant Biol* **27**, 25-47.

Wilkie, N. M., Eglin, R. P., Sanders, P. G., and Clements, J. B. (1980). The association of herpes simplex virus with squamous carcinoma of the cervix, and studies of the virus thymidine kinase gene. *Proc R Soc Lond B Biol Sci* **210**(1180), 411-21.

Wilson, D. S., and Szostak, J. W. (1999). In vitro selection of functional nucleic acids. *Annu Rev Biochem* **68**, 611-47.

Winter, J., Kunath, M., Roepcke, S., Krause, S., Schneider, R., and Schweiger, S. (2007). Alternative polyadenylation signals and promoters act in concert to control tissue-specific expression of the Opitz Syndrome gene MID1. *BMC Mol Biol* **8**, 105.

Yamauchi, Y., Kiriyama, K., Kubota, N., Kimura, H., Usukura, J., and Nishiyama, Y. (2008). The UL14 tegument protein of herpes simplex virus type 1 is required for efficient nuclear transport of the alpha transinducing factor VP16 and viral capsids. *J Virol* **82**(3), 1094-106.

Zhang, Z., and Gilmour, D. S. (2006). Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell* **21**(1), 65-74.

Zhao, G., Shi, L., Qiu, D., Hu, H., and Kao, P. N. (2005). NF45/ILF2 tissue expression, promoter analysis, and interleukin-2 transactivating function. *Exp Cell Res* **305**(2), 312-23.

Zhao, J., Hyman, L., and Moore, C. (1999). Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* **63**(2), 405-45.

Zhao, W., and Manley, J. L. (1998). Deregulation of poly(A) polymerase interferes with cell growth. *Mol Cell Biol* **18**(9), 5010-20.

Zhou, Z. H., Dougherty, M., Jakana, J., He, J., Rixon, F. J., and Chiu, W. (2000). Seeing the herpesvirus capsid at 8.5 A. *Science* **288**(5467), 877-80.